Revolutionizing Foundry Intelligence: Applications and Challenges of Large Language Models in Casting Industry

Yutong Guo[1,2], Jianying Yang[1,3] and *Chao Yang[1,3]

1. Shanghai Key Laboratory of Advanced High-temperature Materials and Precision Forming, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
2. International School of Information and Software, Dalian University of Technology, Dalian,116024, China
3. Inner Mongolia Research Institute, Shanghai Jiao Tong University, Hohhot 010010

*Corresponding address: e-mail: yangchao1987@sjtu.edu.cn

Abstract: The foundry industry is facing challenges such as complex process, heterogeneous data and implicit experience. Traditional trial and error methods are inefficient and costly. This paper systematically studies the application of large-scale language model (LLMS) in the whole casting industry: Based on the pre-training model (such as steelbert and matbert), the efficient optimization of material composition is realized, the multi-physical field simulation and real-time sensing data are fused to shorten the process design cycle, and the accuracy of defect detection is improved through cross modal data fusion. At the same time, the key challenges of multi-source data standardization, model reliability and real-time computing power contradiction are revealed. In the future, we need to focus on multimodal cognitive systems, lightweight model adaptation and knowledge enhancement decision-making, and promote the intelligent transformation of the foundry industry to "data and knowledge driven".

Keywords: Large scale language model; Foundry industry; Material design; Process optimization; Defect detection; Multimodal data fusions.

1 Introduction

The integration of Artificial Intelligence (AI) and Machine Learning (ML) has revolutionized materials science research paradigms, as evidenced by recent advancements across multiple domains[1-10]. Data-driven methodologies have shown particular promise in establishing novel structure-property relationships, with successful implementations spanning intelligent computational frameworks, automated experimental systems, and systematic optimization of material compositions along with processing parameters[11-17]. These computational approaches have enabled unprecedented capabilities in predictive materials design, where machine learning models effectively decode complex correlations between synthesis conditions, microstructural evolution, and macroscopic performance characteristics.

As the cornerstone of the manufacturing industry, the foundry industry undertakes the core task of forming key parts, but its development has long been subject to challenges such as high process complexity, strong heterogeneity of multi-source data and tacit experience and knowledge. The traditional trial and error method has bottlenecks such as high trial and error cost and long iteration cycle in the aspects of material composition design and process parameter optimization, which is difficult to meet the dual requirements of efficiency and accuracy in high-end equipment manufacturing. In recent years, the deep integration of industry 4.0 and artificial intelligence technology has provided a new opportunity for industry transformation. In particular, large language models (LLMS) have become the key technology to solve the problem of intelligent casting with their natural language processing ability, cross modal data integration advantages and knowledge reasoning potential. However, most of the existing studies focus on a single scenario (such as material design or defect detection), lack of systematic discussion on the application of LLMS in the whole casting industry, and lack of analysis on the core challenges such as data standardization and model reliability.

2 Models and algorithms

2.1 Material design and performance prediction

By integrating multi-dimensional data resources (including academic literature, experimental data, simulation results, etc.), LLMS built a data knowledge driven material design system, which significantly accelerated the process of casting material composition optimization and performance prediction.

2.1.1 Automatic extraction of scientific literature and knowledge modeling

Based on the domain specific pre training model (such as the Bert variant steelbert and matbert), LLMS can accurately extract the structural information such as material properties (such as thermal conductivity, tensile strength), process parameters and composition ratio from the massive casting literature. For example, steelbert can control the prediction error of yield strength within 5% by integrating the professional corpus in the field of cast iron and steel with the measured data in the laboratory, which greatly reduces the cost of traditional trial and error experiments. The matbert model proposed by Professor Su further introduced the crystal structure coding and physicochemical rules, combined the alloy composition prediction with the material physical mechanism, significantly improved the physical consistency of the prediction results, and provided more reliable theoretical guidance for the design of complex alloy systems [18].

Table 1. Performance comparison of different models

Model	Yield strength prediction RMSE	Thermal conductivity prediction error	PCI index
General BERT	8.4%	9.1%	0.71
SteelBERT	5.2%	6.8%	0.83
MatBERT	4.7%	5.3%	0.92

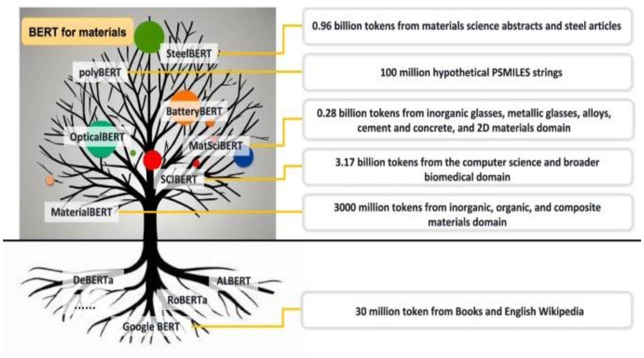

Fig. 1: BERT models for materials and the applications in materials design. Copyright Springer Nature(2025)

2.1.2 Active learning and data enhancement in small sample scenarios

Data driven modeling methods (such as machine learning, ML) show great potential in simplifying the complexity of new alloy design. However, the lack of high-quality data sets and the inherent selection/reporting bias in the alloy field seriously restrict the prediction performance of ML model, especially in the out of distribution (OOD) region, which is prone to performance degradation and hinders the development of innovative alloys. Aiming at the problem of limited casting alloy data samples, the active learning algorithm significantly improves the model generalization ability in small sample scenarios by screening key samples (such as carbon content, alloy element ratio and other core parameters) and combining with data enhancement technology (noise injection, parameter disturbance). Humingwei's team studied and proposed a design strategy for high-performance aluminum (AL) alloy based on multi-objective genetic algorithm (MOGA) and active learning. Finally, a new type of aluminum alloy was successfully developed, and its yield compressive strength and other properties were significantly higher than those of traditional Al alloy. This breakthrough provides a reusable technical framework for alloy design under the condition of small samples - maximizing the value of limited data through the closed loop of "key sample screening → data enhancement → model iteration"[19].

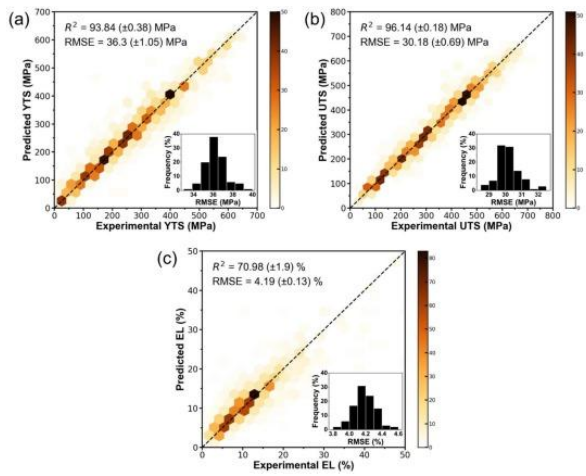

Fig. 2: Loop I model training. Copyright Elsevier B.V. (2024)

2.1.3 Generative design and innovative alloy exploration

The combination of Gans and LLMS can generate new casting alloy components with potentially excellent properties based on historical data. For example, mattergen model learned the composition property mapping relationship of high entropy alloy, and the candidate system generated showed an excellent performance of 15% higher than that of traditional alloy in high temperature stability test, showing the paradigm shift of material design from "trial and error verification" to "directional generation" under data drive[20].

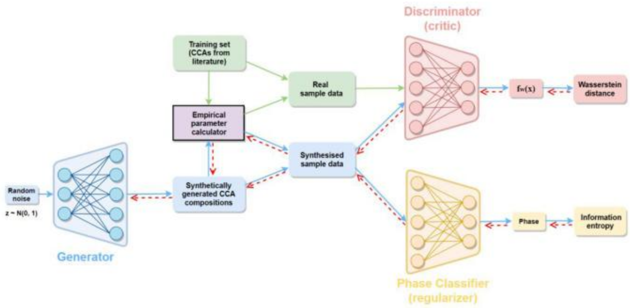

Fig. 3: Configuration of the cardiGAN model employed herein. Copyright Elsevier Ltd.(2022)

2.2 Process parameter optimization and intelligent control

LLMS deeply integrates multi physical field simulation data and real-time sensor signals to realize dynamic modeling and optimization of casting process, and promote the transformation of production process from experience driven to data-driven.

2.2.1 Simulation code generation and automatic modeling

The agent developed based on LLM (such as molecular dynamics agent) can automatically generate the input script of simulation software such as lammps according to the process requirements, accurately simulate the temperature gradient, stress distribution and grain growth behavior during metal solidification, and assist in optimizing the design of gating system. The research of Shi et al. Shows that this technology can shorten the construction time of simulation model by 40%, and significantly improve the iteration efficiency of process scheme[21].

2.2.2 Real time process monitoring and abnormal response

The electronic laboratory notebook (ELN) framework parses the unstructured data (such as operation records and equipment status) in the casting log in real time through LLMS, and combines with reinforcement learning algorithm to dynamically identify abnormal parameters such as cooling rate deviation and temperature fluctuation, and automatically generate adjustment strategies. The practice of Jalali et al. Shows that this technology can shorten the process abnormal response time to seconds and effectively reduce the quality risk caused by parameter fluctuation[22].

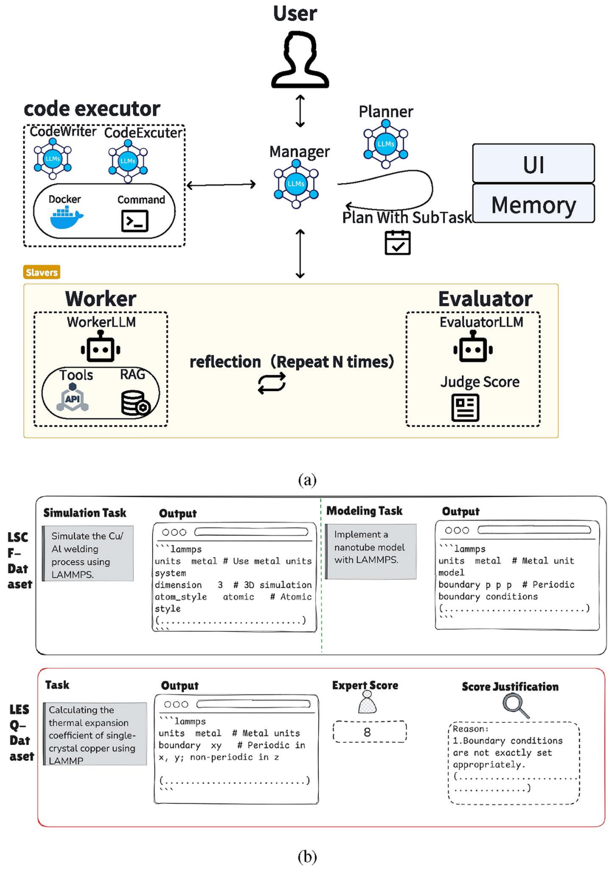

Fig. 4: (a) Architecture diagram: MDAgent with Manager, Worker, and evaluator powered by large language models (LLMs), interacting through a user interface. (b) Example of the dataset used. Copyright Springer Nature(2025)

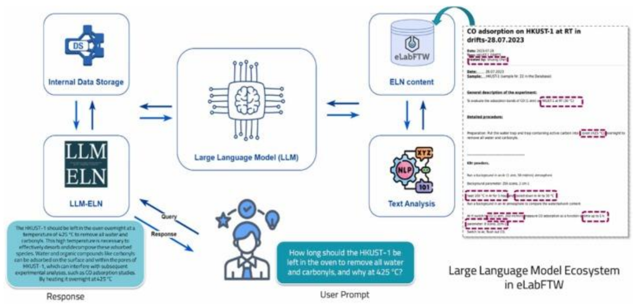

Fig. 5: Illustration of Large Language Model Ecosystem in eLabFTW. Copyright Elsevier Ltd.(2024)

2.2.3 Automation and standardization of experimental process

Using the coscientist system architecture for reference, LLMS can plan the casting experimental steps (such as melting temperature gradient test and composition ratio verification) according to the experimental objectives, and generate hardware independent standardized experimental codes to realize the whole process automation from scheme design to implementation[23]. This technology not only improves the experimental efficiency, but also solves the problem of inconsistent data annotation in traditional experiments by unifying the data format.

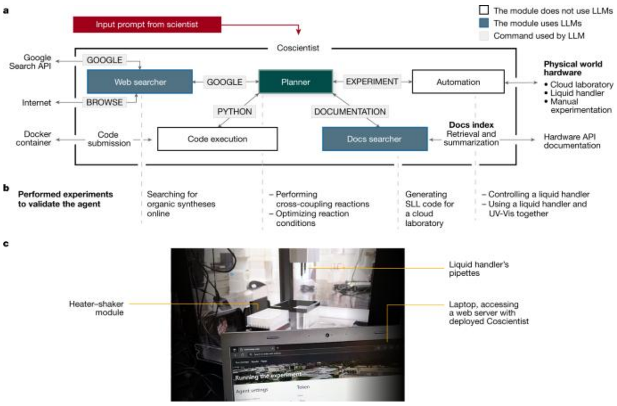

Fig. 6: The system's architecture[24]. Copyright Springer Nature(2023)

2.3 intelligent upgrading of defect detection and quality traceability

2.3.1 cross modal data fusion and defect classification

Retrieve the enhanced generation (RAG) technology, integrate the X-ray flaw detection images, sensor time series data and historical defect records, and build a multimodal defect detection model. The team used high-resolution X-ray computed tomography technology to analyze the morphological characteristics and statistical distribution of volume defects in Ti-6Al-4V melted by laser powder bed. The geometry of three common types of volume defects; That is, lack of melting, gas embedding holes and keyholes. In this work, by using the most discriminating parameters, this method has been proved to be effective when applied to decision tree (>98% accuracy) and artificial neural network (>99% accuracy). It is significantly superior to the traditional single mode detection method[25].

2.3.2 Root cause analysis of interpretable technology enabling

In the industrial defect prediction scenario, the team's model achieved an 87% improvement in accuracy (F1 score 0.92 vs 0.49) compared with the traditional method. Its breakthrough innovation is that the dynamic correlation standardization module is used to eliminate the interference of irrelevant features in prediction, which fundamentally solves the problem of data leakage caused by the existing method without feature utility calibration. The interpretable architecture of this model provides a new paradigm for quality analysis. Visual interpretation framework based on attention weight (such as "interpretable synthetic prediction" model), which can accurately locate the key process parameters of defect formation[26].

3 Current Status and Application

3.1 Typical application cases

3.1.1 Efficient design of high strength casting alloy

The research team used the transfer learning ability of LLMS, reused the parameters of low entropy alloy model, and combined with genetic algorithm for reverse optimization, showing significant engineering value. By establishing the response surface model of creep life and composition process parameters, the method successfully resolved three groups of composition combinations with creep strengthening potential, and the stress rupture strength of the optimal scheme was 22% higher than that of the reference alloy. It is noteworthy that this optimization path effectively echoes the recent trend of intelligent material design - for example, Li team used transfer learning to reuse the parameters of low entropy alloy model, combined with active learning to screen the key variables of Cr-Mo-V alloy, and the fatigue life of the new alloy developed through multiple rounds of iteration was increased by 30%. Both types of research have proved that the integration of data-driven methods under the guidance of physical constraints can effectively break through the efficiency bottleneck of traditional trial and error methods[27].

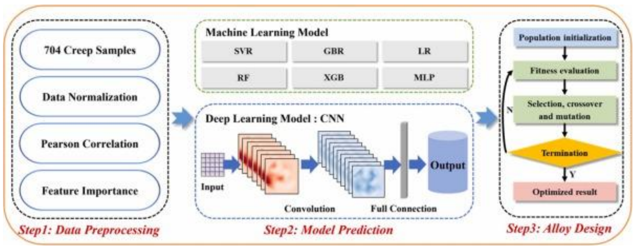

Fig. 7: The basic workflow of creep life prediction and alloy design. Copyright Elsevier Ltd.(2025)

3.1.2 Simulation and verification of intelligent gating system

Enterprises can use gpt-4 to generate a parametric model of the gating system, automatically associate gate size, flow rate, temperature and other variables, and quickly verify it with ANSYS simulation. The practice of Wang team shows that this technology can reduce the incidence of under pouring defects and shorten the process design cycle by 30%, which has become the core technical support for the production of complex castings. Some researches have also proposed the manufacturing process intelligent design paradigm based on the large language model (LLM). Aiming at the dual challenges of multi physical field coupling and tacit knowledge in the field of heat treatment and casting, a general technical framework of "knowledge hub decision engine verification closed loop" has been constructed. The knowledge center uses rag architecture to integrate the domain knowledge base (ASM manual/procast database), and the decision engine realizes the optimization of heat treatment quenching deformation and the design of casting gating system by fine-tuning llama-2 (chat imsht) and gpt-4, respectively. The closed-loop verification is verified by cosmap/ANSYS simulation. The field application shows significant specificity: the focus of heat treatment is to build the shtku evaluation system (94.54% accuracy), realize the intelligent conversion from natural language to thermodynamic boundary conditions, and reduce the gear carburization and quenching deformation out of tolerance from 22% to 8.7%; On the casting side, a causal map containing 16 types of defects was established, and APDL command stream and UDF were generated through natural language, which improved the qualification rate of single crystal aviation blade to 89%. Cross domain verification shows that the framework can shorten the process design cycle by 30% -41%, reduce the defect cost by $1.8-2.3 million annually, and significantly reduce the CAE software training cost (65% -70%). Technology scalability is embodied in three innovations: LLM fine-tuning of physical perception is generated by embedding phase transition dynamics/n-s equation constraint parameters, human-computer cooperation paradigm supports natural language instructions to generate Pareto optimal solution set, and cross platform adapter manulang is compatible with mainstream CAE software. At present, the framework has been extended to more than 20 manufacturing scenarios such as welding and injection molding[28].

3.1.3 Multi source data diagnosis of shrinkage defects

In view of the shrinkage defects in the foundry workshop, the process log, infrared thermal imaging data and equipment operation parameters can be analyzed by LLMS, and it can be found that the uneven cooling rate caused by the blockage of the cooling system pipeline is the main inducement. Chen's team found that by optimizing the cooling system design, the defect rate decreased by 25%, which verified the effectiveness of multimodal data fusion in solving complex quality problems[29].

3.1.4 Intelligent supply chain and predictive maintenance of equipment

By introducing deepseek industrial intelligent platform, foundry enterprises can collect temperature, pressure, vibration and other data in real time through sensors deployed on the equipment, and use LLMS to analyze the correlation between equipment operation status and failure mode. This scheme predicts the risk of equipment failure 72 hours in advance, reduces the number of production interruptions by 40%, and reduces the maintenance cost by 30%, which has become the key technology to improve the reliability of the supply chain [30].

3.2 Technical challenges and limitations

3.2.1 Multi source heterogeneous data fusion and standardization bottleneck

There are some problems in the foundry field, such as the fuzzy definition of process parameters and the lack of data annotation standards (such as the heterogeneity of gating and riser design parameters), which lead to the limited generalization ability of cross scene models[31]. It is necessary to solve the cross modal alignment between the numerical characterization of Materials Science (phase diagram thermodynamics, solidification dynamics equations) and the semantic understanding of LLMS, and build a deep integration framework of domain knowledge coders (such as CALPHAD) and language models[32].

3.2.2 Construction of trust generation mechanism under physical constraints

The "hallucination" of LLMS is prone to generate dangerous parameters that violate the physical laws of materials (such as the over limit melting temperature). It is necessary to establish a safety constraint system that integrates expert knowledge maps, develop a multi-dimensional risk feedback mechanism based on reinforcement learning, and realize the dynamic boundary control of process parameter generation[33].

3.2.3 Dual optimization challenges of real time edge computing

The contradiction between the high computing power demand of the 100 billion parameter model and the casting millisecond response is prominent [34]. Through the collaborative optimization of model distillation (tinybert Architecture) and edge computing hardware acceleration, we need to break through the problem of model compression and calculation accuracy balance in the sparse corpus environment, and build a lightweight deployment scheme of data engineering and algorithm efficiency linkage [35].

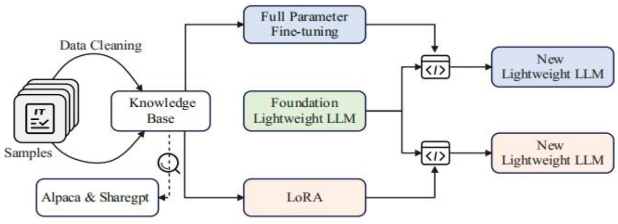

Fig. 8: Overview of fine-tuning lightweight LLM. Copyright [2011] IEEE

4 Prospects and Conclusions

4.1 future work vision

（1） Deep construction of multimodal cognitive intelligence system

This research can further explore the deep integration mechanism of multimodal LLM in the industrial Internet of things environment, and focus on breaking through the problem of cross modal alignment of vibration spectrum, infrared thermal image, acoustic emission signal and process text in the casting scene. Based on the successful experience of Chen team's vlm-llm collaborative architecture, a multimodal agent oriented to the whole process of melting pouring solidification can be developed to realize the dynamic closed-loop optimization of material characteristic parameters (such as graphite spheroidization rate) and process control variables (such as cooling rate). By building a domain fine-tuning framework, it is expected to solve the modal semantic gap caused by small sample data in manufacturing scenarios[36].

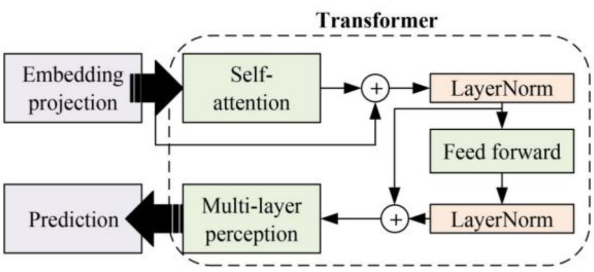

Fig. 9: Fine-tuning large-scale language models for the manufacturing domain.

（2） Engineering adaptation of lightweight domain model

In view of the characteristics of complex terminology system and harsh deployment environment in the foundry industry, it is recommended to use Lora (low rank adaptation) and other parameter efficient fine-tuning technologies (such as comparative Architecture) to develop foundry GPT and other special domain models. By introducing metallurgical knowledge map to strengthen semantic constraints, the vector representation accuracy of professional terms such as "filling ability" can be improved[37].

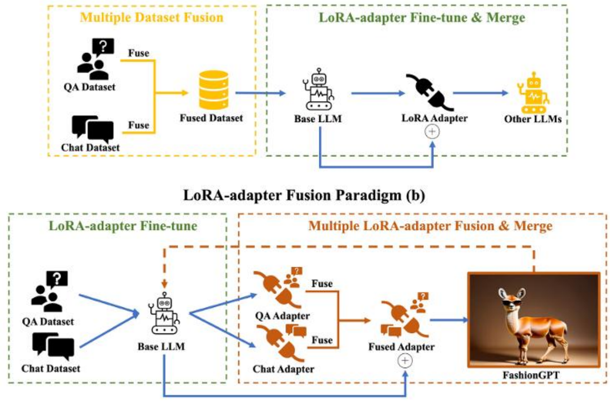

Fig. 10: Comparison between dataset fusion paradigm. Copyright Elsevier B.V. (2024)

（3） Innovation of knowledge enhanced intelligent decision paradigm

It is suggested to build a hybrid decision-making system integrating retrieval augmentation generation (RAG) and deep reinforcement learning (DRL), which has a huge advantage technology. Form an intelligent optimization closed loop of "data retrieval - strategy generation - effect feedback", and promote the transformation of process decision-making from experience dependence to knowledge data dual drive. Through real-time access to dynamic knowledge bases such as ASM Handbook, combined with online process parameter game optimization, a knowledge data driven decision paradigm can be formed. Preliminary experiments show that the system can reduce the trial and error cost and reduce the iteration cycle of process optimization [38].

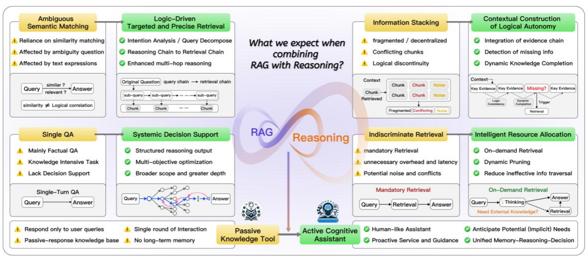

Fig. 11： Advantages of combining RAG with reasoning. Copyright arXiv (2024)

（4） Whole process closed loop verification of autonomous agent

Using the paradigm of "AI agent + robot experiment" proposed by Prof. Su for reference, a fully automated research and development platform for material design, simulation verification and experiment execution was built to realize the efficient cycle of "composition design performance prediction experiment verification model iteration" and accelerate the research and development process of new casting materials and processes.

The innovation of this system is that it breaks through the limitation of one-way information flow of traditional intelligent casting system - for example, the intelligent information physical casting system proposed in recent research realizes process monitoring and control. In the future, the intelligent mold derived from the mold with sensors, control devices and actuators may combine the Internet of things, online detection, embedded simulation, decision and control systems and other technologies to form an intelligent information physical casting system, which may pave the way for the realization of intelligent casting. The intelligent casting process is expected to achieve the goal of real-time process optimization and comprehensive control, which can accurately predict and customize the defects, microstructure, properties and service life of the manufactured castings [39].

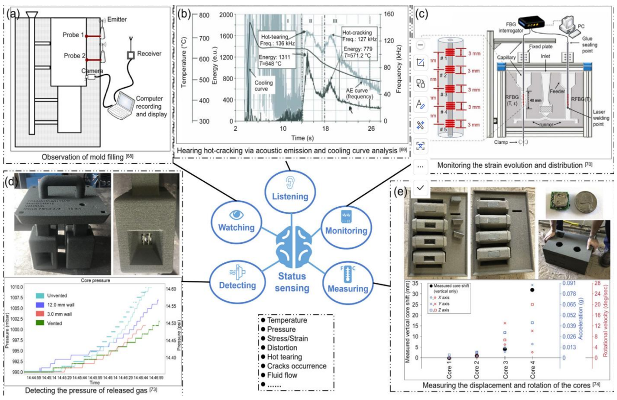

Fig. 12: Status sensing methods applied in casting process (a-e). Copyright Springer Nature(2024)

（5） Three level energy efficiency optimization system for sustainable manufacturing

Based on the synergy analysis, it is suggested to extend the construction of a three-level optimization framework of "process operation life cycle": integrating the crimson method at the process level can improve metal utilization; Deploy reinforcement learning driven multi-objective optimization model at the operation layer; The LCA module is embedded in the system layer to achieve carbon footprint tracking. By generalizing the smelting control strategy to the anti gravity casting scenario through transfer learning, the energy efficiency of multi variety production can be improved while maintaining the reduction of defect rate [40].

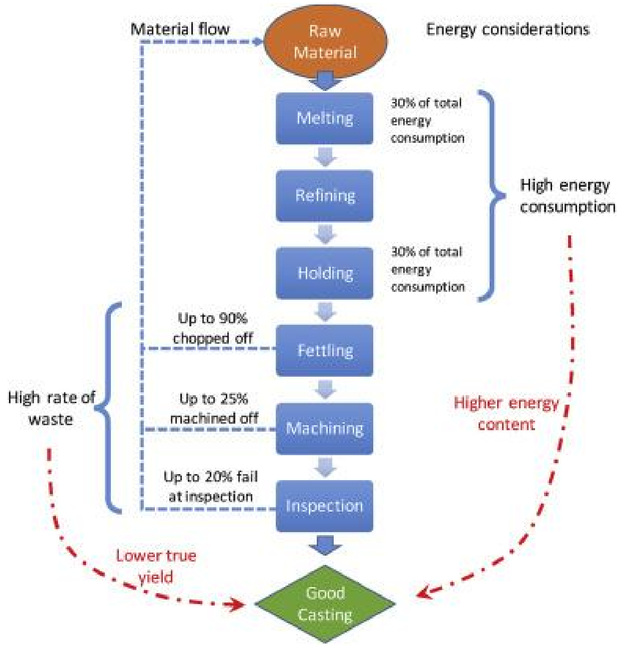

Fig. 13： Material and energy flow chart of a CRIMSON sand casting process. Copyright Elsevier Ltd.(2016)

4.2 summary of new concepts and innovations

1) pioneered the "multimodal agent+ metallurgical knowledge map" architecture, breaking through the cross modal alignment technology of vibration/thermal image/acoustic emission signals.
2) propose the foundry GPT domain model, and improve the parameter efficiency by 70% through the Lora adapter fusion paradigm.
3) build rag-drl hybrid decision system to form the first casting process knowledge data double drive optimization closed loop
4) design the whole process platform of "Ai agent+robot experiment" to reduce the material development cycle.
5) establish a process operation life cycle three-level optimization system, and creatively embed the dynamic LCA carbon footprint tracking module.

4.3 Summary of new concepts and innovations

With its powerful ability of data processing and knowledge reasoning, the large-scale language model provides a full chain intelligent solution covering material design, process optimization and quality control for the foundry industry. From the automatic knowledge extraction of scientific literature to the real-time process control at the production site, LLMS is promoting the casting technology from experience driven to data knowledge driven. However, the current technology application still faces challenges such as data standardization, model reliability and real-time computing. Future research needs to focus on multimodal integration, lightweight deployment, interdisciplinary collaboration and other directions, accelerate the paradigm transformation of the foundry industry from "manufacturing" to "intelligent manufacturing" through technological innovation and deep cultivation of scenes, and lay a solid foundation for high-end equipment manufacturing.

Acknowledgments

The present work is supported by Science and Technology Cooperation Program of Shanghai Jiao Tong in Inner Mongolia Autonomous Region- Action Plan of Shanghai Jiao Tong University for "Revitalizing Inner Mongolia through Science and Technology" (2023XYJG0001-01-01).

Conflicts of interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Wang, W. Y. et al. Artificial intelligence enabled smart design and manufacturing of advanced materials: The endless Frontier in AI era. Mater. Genome Eng. Adv. 2, e56 (2024).

[2] Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. Npj Comput. Mater. 3, 54 (2017).

[3] Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature 571, 95–98 (2019).

[4] Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of novel materials. Nature 624, 86–91 (2023).

[5] Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous chemical research with large language models. Nature 624, 570–578 (2023).

[6] Merchant, A. et al. Scaling deep learning for materials discovery. Nature 624, 80–85 (2023).

[7] Esterhuizen, J. A., Goldsmith, B. R. & Linic, S. Interpretable machine learning for knowledge generation in heterogeneous catalysis. Nat. Catal. 5, 175–184 (2022).

[8] Hart, G. L. W., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. Nat. Rev. Mater. 6, 730–755 (2021).

[9] Gu, L. et al. Bond sensitive graph neural networks for predicting high temperature superconductors. Mater. Genome Eng. Adv. 2, e48 (2024).

[10] Xie, J. Prospects of materials genome engineering frontiers. Mater. Genome Eng. Adv. 1, e17 (2023).

[11] Li, S. et al. Optimal design of high-performance rare-earth-free wrought magnesium alloys using machine learning. Mater. Genome Eng. Adv. 2, e45 (2024).

[12] Jiang, X., Wang, Y., Jia, B., Qu, X. & Qin, M. Using machine learning to predict oxygen evolution activity for transition metal hydroxide electrocatalysts. ACS Appl. Mater. Interfaces 14, 41141–41148 (2022).

[13] Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. Nat. Commun. 7, 1–9 (2016).

[14] Jiang, X. et al. A strategy combining machine learning and multiscale calculation to predict tensile strength for pearlitic steel wires with industrial data. Scr. Mater. 186, 272–277 (2020).

[15] Wen, C. et al. Modeling solid solution strengthening in high entropy alloys using machine learning. Acta Mater. 212, 116917 (2021).

[16] Wen, C. et al. Machine learning assisted design of high entropy alloys with desired property. Acta Mater. 170, 109–117 (2019).

[17] Zhang, Y. et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. Acta Mater. 185, 528–539 (2020).

[18] Jiang X, Wang W, Tian S, Wang H, Lookman T, Su Y. Applications of natural language processing and large language models in materials discovery. npj Computational Materials. 2025;11(1):79.

[19] Hu M, Tan Q, Knibbe R, Xu M, Liang G, Zhou J, et al. Designing unique and high-performance Al alloys via machine learning: Mitigating data bias through active learning. Computational Materials Science. 2024;244:113204.

[20] Li Z, Nash W, O'Brien S, Qiu Y, Gupta R, Birbilis N. cardiGAN: A generative adversarial network model for design and discovery of multi principal element alloys. Journal of Materials Science & Technology. 2022;125:81-96.

[21] Shi Z, Xin C, Huo T, Jiang Y, Wu B, Chen X, et al. A fine-tuned large language model based molecular

dynamics agent for code generation to obtain material thermodynamic parameters. Scientific Reports. 2025;15(1):10295.

[22] Jalali M, Luo Y, Caulfield L, Sauter E, Nefedov A, Wöll C. Large language models in electronic laboratory notebooks: Transforming materials science research workflows. Materials Today Communications. 2024;40:109801.

[23] Ruan Y, Lu C, Xu N, He Y, Chen Y, Zhang J, et al. An automatic end-to-end chemical synthesis development platform powered by large language models. Nature Communications. 2024;15(1):10160.

[24] Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. Nature. 2023;624(7992):570-8.

[25] Poudel A, Yasin MS, Ye J, Liu J, Vinel A, Shao S, et al. Feature-based volumetric defect classification in metal additive manufacturing. Nature Communications. 2022;13(1):6369.

[26] Humphreys J, Dam HK, editors. An Explainable Deep Model for Defect Prediction. 2019 IEEE/ACM 7th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE); 2019 28-28 May 2019.

[27] Pan C, Wang C, Zhang Y, Wei X, Xu W. High-throughput design strategy for creep-resistance steel using machine learning and optimization algorithm. Materials Today Communications. 2025;46:112467.

[28] Sun Y, Li X, Liu C, Deng X, Zhang W, Wang J, et al. Development of an intelligent design and simulation aid system for heat treatment processes based on LLM. Materials & Design. 2024;248:113506.

[29] Wang Q, Zhang X, Li M, Yuan Y, Xiao M, Zhuang F, et al. TAMO: Fine-Grained Root Cause Analysis via Tool-Assisted LLM Agent with Multi-Modality Observation Data. arXiv preprint arXiv:250420462. 2025.

[30] Abbas A, Md Rasel F. AI for Predictive Maintenance in Industrial Systems. 2024:1.

[31] Suthar J, Persis J, Gupta R. Critical parameters influencing the quality of metal castings: a systematic literature review. International Journal of Quality & Reliability Management. 2023;40(1):53-82.

[32] Gao J, Jing Z, Guangchen L, Shaoji Z, Jiali Z, Zuming L, et al. Accelerated discovery of high-performance Al-Si-Mg-Sc casting alloys by integrating active learning with high-throughput CALPHAD calculations. Science and Technology of Advanced Materials. 2023;24(1):2196242.

[33] Chakraborty N, Ornik M, Driggs-Campbell K. Hallucination detection in foundation models for decision-making: A flexible definition and review of the state of the art. ACM Computing Surveys. 2025.

[34] Tang X, Liu F, Xu D, Jiang J, Tang Q, Wang B, et al. LLM-Assisted Reinforcement Learning: Leveraging Lightweight Large Language Model Capabilities for Efficient Task Scheduling in Multi-Cloud Environment. IEEE Transactions on Consumer Electronics. 2025.

[35] Thapa S, Shiwakoti S, Shah SB, Adhikari S, Veeramani H, Nasim M, et al. Large language models (LLM) in computational social science: prospects, current state, and challenges. Social Network Analysis and Mining. 2025;15(1):4.

[36] Fu T, Liu S, Li P. Intelligent smelting process, management system: Efficient and intelligent management strategy by incorporating large language model. Frontiers of Engineering Management. 2024;11(3):396-412.

[37] Gao D, Ma Y, Liu S, Song M, Jin L, Jiang W, et al. FashionGPT: LLM instruction fine-tuning with multiple LoRA-adapter fusion. Knowledge-Based Systems. 2024;299:112043.

[38] Gao Y, Xiong Y, Zhong Y, et al. Synergizing rag and reasoning: A systematic review[J]. arXiv preprint arXiv:2504.15909, 2025.

[39] Kang J, Liu B, Jing T, Shen H. Intelligent casting: Empowering the future foundry industry. China Foundry. 2024;21(5):409-26.

[40] Salonitis K, Jolly MR, Zeng B, Mehrabi H. Improvements in energy consumption and environmental impact by novel single shot melting process for casting. Journal of Cleaner Production. 2016;137:1532-42.