研究论文

金属晶界性质及相应力学性能的数智化研究进展

李昕,张志伟,高旺*,蒋青

摘要 晶界是金属材料中的常见缺陷之一,严重影响金属的力学性能。针对工程应用对金属材料高强度与抗损伤性能的迫切需求,以晶界性质及相应力学性能为例,综述了金属材料微观结构与力学性能的数智化研究进展。阐述了“高通量计算—理论模型—机器学习与大模型”的闭环研究范式:高通量计算平台通过多尺度计算批量生成材料数据,形成数据库;在此基础上,结合材料经典热动力学模型与金属键合性质模型,提炼具有物理意义的描述符,构建理论模型,建立晶界微观结构与宏观力学性能的定量构效关系;机器学习可用于材料性质的快速预测,还可融合描述符与理论模型解决其“黑箱”问题,进而指导理论计算和实验;大模型可基于海量的材料科学语料进行训练,展现出强大的跨领域知识整合与生成能力,可作为高效的智能研究助理辅助材料发现与设计。

关键词 高通量计算;材料数据库;理论模型;描述符;机器学习;大模型;金属晶界;力学性能

金属晶界作为多晶材料中典型的界面缺陷,其结构和性质与材料的强化脆化密切相关,是调控材料宏观力学性能的关键微观结构之一。例如,高强钢的氢致开裂与合金的应力腐蚀均与晶界化学成分及结构密切相关,晶界的高温稳定性关乎抗蠕变强度与抗热疲劳断裂能力;辐照条件下,晶界对辐照点缺陷的捕获与演化强烈影响着其抗辐照脆化性能。这些关键材料的力学性能,本质上取决于其微观结构。因此,从原子尺度揭示晶界结构、成分与能量对其力学性能的影响机制,已成为突破高性能金属材料设计瓶颈的必要前提之一。通过系统探究材料晶界的微观结构及其与各类缺陷的相互作用,建立从“缺陷调控”到“性能涌现”的构效关系,为实现重大工程装备的原创性设计与材料自主研发提供科学依据。

晶界强化(细晶强化)是金属的强化机制之一,通过增加晶界密度阻碍位错运动,遵循 Hall−Petch 关系,实现强度的提升;当第二相颗粒在晶界处或晶内析出时,可产生显著的析出强化,其强化效果取决于第二相粒子尺寸、分布及其与晶界等缺陷的相互作用。此外,晶界脆化源于溶质元素在晶界偏聚,大幅降低晶界结合能;氢脆是氢原子在位错、晶界等缺陷处富集,促进微裂纹形核与扩展;在辐照环境中,点缺陷聚集形成的溶质团簇可能向晶界处偏析或沉淀,引发辐照脆化。因此,对晶界行为的精准调控,包括晶界的成分、结构、类型、密度、分布以及晶界与其他类型缺陷或环境因素的相互作用,成为实现材料强韧化、预防金属灾难性破坏的关键所在。系统揭示晶界性质与宏观力学性能之间的定量构效关系,是从本质

吉林大学材料科学与工程学院,汽车材料教育部重点实验室,长春 130022

收稿日期:2025−12−08;修回日期:2025−12−29
基金项目:国家自然科学基金项目(22173034,11974128,52130101)
作者简介:李昕,博士研究生,研究方向为金属计算材料学,电子信箱:lx21@mails.jlu.edu.cn;高旺(通信作者),教授,研究方向为金属计算材料学,电子信箱:wgao@jlu.edu.cn
引用格式:李昕,张志伟,高旺,等. 金属晶界性质及相应力学性能的数智化研究进展[J]. 科技导报,2026,44(18):137−151;doi:10.3981/j.issn.1000-7857.2025.12.00041

上理解复杂材料强化或失效机理,突破现有工程设计极限,实现从“经验设计”到“预测设计”转变的关键。这一针对晶界的研究,实质上是更广泛的“微观缺陷—宏观性能”构效关系研究在界面科学中的核心体现。这一构效关系指导着经典的“成分—工艺—结构—性能”4 要素研究闭环:通过精确调控成分与优化制备工艺,获得具备更高强度和韧性、有效抑制损伤萌生、延缓损伤演化的理想材料,最终实现材料在复杂工况下长寿命、高可靠服役性能的既定目标。

为深刻揭示金属材料“微观缺陷—宏观性能”之间的内在构效关系,实现对力学性能的精准预测与理性调控,理论研究正在向“数智化”迈进,即计算模拟、数据科学和人工智能技术与传统材料理论的深度融合,形成以计算支撑、数据驱动、理论引导与人工智能加速为核心的新研究范式,如图 1 所示[1]。具体而言,“数智化研究”以集成先进计算工具为基础,通过高通量计算产生大量数据,构建材料数据库,从金属键合性质与经典材料热动力学等理论模型出发,量化材料特征,构建材料基因(描述符),结合机器学习方法与大模型,实现对材料微观性质与宏观性能的精准预测。因此,“理论研究数智化”旨在建立从材料基因到微观机制再到宏观行为的完整映射关系,为突破传统理论局限、设计新型强韧化材料提供全新思路。晶界作为多晶材料中普遍存在的面缺陷,其微观结构特征与宏观力学性能之间的构效关系研究,为理解材料数智化研究进展提供了理想范例。因此,以晶界性质及相关力学性能为重点,总结数值计算、解析模型、机器学习模型及大模型的发展,深入探讨不同方法在揭示金属强化与脆化机制中的技术特点与独特价值,对这一数智化研究体系的进展予以综述。

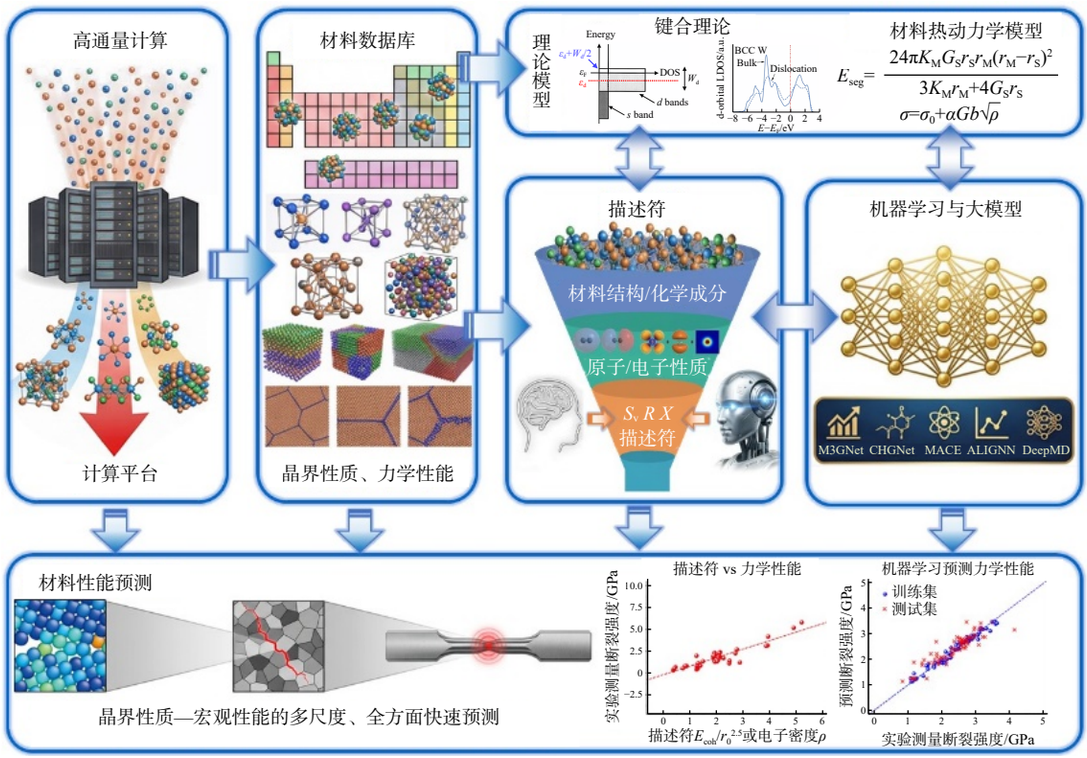


图 1 金属材料晶界性质及力学性能数智化研究示意

1 高性能计算平台及数据库

高性能计算平台与多尺度模拟方法是实现高通量计算、贯通跨尺度研究的数值计算基础,致力于将材料的成分、微观结构与宏观性能进行定量关联,为揭示金属强化和脆化的内在物理机制提供了“虚拟实验室”,支撑智能计算新范式。当前,这一领域已从依赖单一尺度的模拟,演进至第一性原理、分子动力学、相场法及晶体塑性有限元等多尺度计算框架集成协同应用,实现了从原子键合、缺陷性质到宏观性能的跨尺度贯通与系统关联。图 2 为高通量计算平台材料数据库。

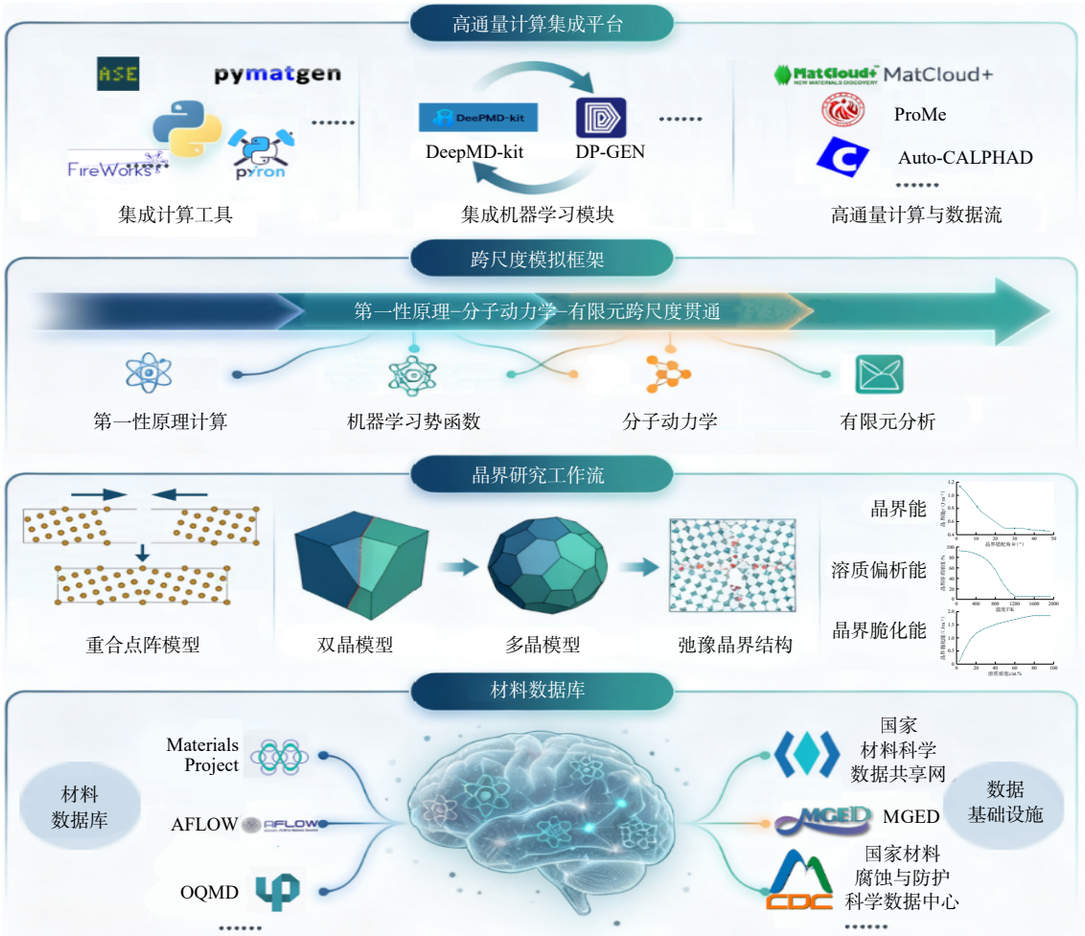


图 2　高通量计算平台与材料数据库示意

1.1　高性能计算平台

晶界的计算研究始于其原子级结构的构建。晶界指同一相内 2 个晶粒之间的界面，最常用的方法是构建双晶模型。该模型基于晶体学参数，明确定义了晶界的 5 个自由度：两侧晶粒的相对取向（3 个欧拉角）和晶界面的法向。通过重合点阵模型可以系统性地生成一系列低 Σ 值的特殊晶界（如 Σ3、Σ5、Σ11 等），这些晶界通常具有较低的界面能和独特的性质[2]。多晶可基于 Voronoi 镶嵌的物理冶金方法构建，即在晶胞中随机布设一系列晶核，随后根据晶核位置对空间进行 Voronoi 划分，每个 Voronoi 多面体即为一个晶粒。通过为每个晶粒赋予随机或特定的晶体取向，即可构建出具有特定晶粒尺寸分布的三维多晶模型[3]。获得稳定的晶界原子结构后，即可计算其关键物理性质，如反映晶界热力学稳定性的晶界能、预测溶质原子在晶界处的富集倾向的偏析能、表征晶界力学强度和脆性的晶界结合能/断裂功，以及反映晶界动力学性质的晶界扩散势垒等，这些性质是理解其宏观行为的基础。使用传统分子动力学或基于机器学习势函数的分子动力学方法以及晶体塑性有限元等方法，可以直接对包含晶界的多晶模型施加拉伸、剪切或压缩载荷；模拟可以直观地揭示点缺陷向晶界偏析，晶界的滑移或迁移，位错在晶界处形核、发射、被吸收或发生反应，以及裂纹在偏析脆化的晶界处萌生和扩展等。这些动态过程反映了材料屈服、加工硬化、韧性和脆性的原子起源。表 1 列举了部分与晶界性质研究相关的软件平台名称、类别、应用范围、特色功能及侧重的应用领域。

Atomistic Simulation Environment（ASE）[5]是一个基于 Python 的开源库，体现了高性能计算平台在工作流自动化、代码无缝集成与高通量计算方面的价值特点。它支持与数十种第一性原理和分子动力学软件进行交互，其中包括 VASP、CASTEP、CP2K、QE 等第一性原理计算软件及 LAMMPS 等分子动力学软

表 1 部分与晶界性质研究相关的软件平台名称、类别、应用范围、特色功能及侧重的应用领域

平台名称	类别	应用范围及特色功能	侧重的领域
Atomsk[3]	晶体结构建模工具	创建复杂超胞、多晶模型、缺陷结构	晶界和位错等常见缺陷结构建模
Aimsgb[4]	专用晶界生成工具	对称与非对称晶界模型的快速构建,基于重合点阵模型(CSL)理论生成特定Σ值的晶界、支持多种晶体结构	晶界能计算、界面偏析研究的结构准备
ASE[5]	原子模拟框架/接口平台	第一性原理与分子动力学模拟的前后处理、工作流自动化,灵活的原子操作、内置优化算法、与机器学习库(如scikit-learn)无缝衔接等	原子结构的高通量结构弛豫、表面/界面等性质的高通量计算
Pymatgen[6]	材料信息学/结构分析库	晶体结构分析、对称性识别、描述符生成、相图计算。GrainBoundaryGenerator 模块(高通量晶界模型生成)、缺陷分析模块、与 Materials Project 数据库和 Fireworks 等工作流对接等	材料基因、高通量筛选、结构特征工程
Pyiron[7]	集成计算平台	高通量建模与结构分析、多尺度模拟工作流管理、数据存储与分析	复杂合金设计、相变模拟、多尺度材料工程
AFLOW[8]	高通量计算框架	自动化的第一性原理计算、材料性质数据库生成。AFLOW内置模块自动界面性质的计算、与AFLOW数据库联动等	无机晶体、金属间化合物、力学性质预测
DeePMD-kit DP-GEN[9-10]	机器学习势函数平台	高精度分子动力学模拟、势函数开发与主动学习。自动化势函数开发与迭代,确保精度的同时加速计算,对于晶界和位错等大尺度缺陷提供加速研究的工具	大体系/长时间尺度模拟、复杂合金与界面动力学
DAMASK[11]	晶体塑性有限元模拟套件	多晶材料塑性变形、织构演化、损伤分析,基于位错密度的物理本构、热力学耦合、多物理场求解	金属塑性成形、疲劳与断裂、微观组织-性能关联
MatCloud+[12]	材料云计算与数据平台	高通量计算、材料数据库、机器学习建模云端一体化、原子结构建模与性质预测模块、与实验数据关联分析	新材料研发、工艺优化等协同平台
ProMe[13]	极端条件物性计算平台	高压、高温、强辐照等极端环境下材料性质模拟。辐照损伤模拟、高压相变预测、极端条件本构关系	核材料、超硬材料、极端环境材料设计
力牧(Limo)[14]	材料力学性能计算软件	极端变形条件下材料微观结构演化与力学响应。计算多晶界面性质、位错性质、相变动力学以及弹性常数计算等	金属材料、合金设计、力学性能模拟及本构关系的建立
Powerbyte[15]	相场模拟软件	相场计算、相变相关模拟、第二相析出、位错生成等结构模拟	铸造与焊接组织控制、热处理工艺优化、增材制造微观组织
GalaxyEDS	多效应仿真计算机辅助工程(CAE)平台	多物理场耦合分析、工艺-结构-性能一体化仿真	特种冶金、特种加工、复合材料设计、器件多物理场分析

件,最新版本的 ASE 已接入图神经网络力场和深度学习力场(DeepMD-kit)。ASE 还接入了国内首个完全拥有自主知识产权的第一性原理计算软件包 ABACUS[16]和国内首个完全基于图形处理器(graphics processing unit, GPU)加速的开源分子动力学软件包 GPUMD[17]。同时,通过将 ASE 与计算集群的任务调度系统结合,进行批量计算。ASE 在计算过程中产生的原子结构、能量和力学性质等数据,可以方便地与 Python 的数据科学生态系统(如 NumPy 和 Pandas 等)对接,直接用于后续的数据分析和可视化,或者通过原子结构和性质构建描述符并输入到机器学习模型中,实现了从计算到分析的无缝数据流水线。

与之深度集成的 Pymatgen 库[6]进一步扩展了其功能边界,提供了强大的晶体结构分析、对称性识别和材料特征生成能力,形成了从计算执行到数据分析的完整工具链。Pymatgen 提供了高效的 GrainBoundary Generator 模块,成为高通量生成重合点阵晶界结构的重要工具。与 ASE 和 Pymatgen 具有相似功能的计算平台还有针对计算材料学开发的集成环境 pyiron,由 Materials Project 团队开发的用于定义、管理和执行科学工作流的代码库 FireWorks等。这些集成平台已在高通量计算工作中大量使用,并广泛应用于研究金属、金属间化合物以及高熵合金等体系的点缺陷和表界面性质。例如,Zheng 等[2]使用 Pymatgen 的 Grain-

BoundaryGenerator 模块,高通量地生成了 40 多种金属的数百个晶界模型,系统性地建立了不同金属的晶界能数据库,并深入分析了晶界能与晶界几何特征之间的构效关系。Li 等[18]利用 Pymatgen 和 Fireworks 探究了 NbMoTaW 难熔多主元合金的缺陷性质,发现了晶界处 Nb 的富集显著提高了界面稳定性,使其表现出比随机固溶体更高的强度。

在传统金属材料的研发中,晶界和位错等晶体缺陷的精确模拟面临着固有的尺度挑战:这些缺陷的核心区域需要量子力学精度来描述其电子和原子结构,而其长程应力场却延伸至数百纳米,远超第一性原理计算的能力范围。这种尺度跨越使得单一的计算方法难以完整描述缺陷行为。DeepMD-kit[9]通过深度神经网络构建的势函数,既保持了第一性原理的精度,又能实现大规模分子动力学模拟,为多尺度计算提供了关键的支撑;而 DP-GEN 平台与现有工具如分子动力学模拟软件和 DeepMD-kit 衔接,通过其自动化的并发学习框架,系统性地解决了机器学习势函数构建中数据生成这一关键瓶颈,为材料计算模拟提供了一个能够自动调度迭代过程、高效利用计算资源并最终生成准确势能面模型的工具[10]。这些框架在保证量子力学计算精度的同时,能够模拟包含完整应力场的大尺度体系,成功应用于 Al-Cu 合金、Fe-Si 合金的界面性质、铝合金的晶界偏析、金属钨的缺陷-晶界相互作用等[19-22]一系列研究,实现了从依赖经验势函数的近似模拟到基于第一性原理精度的、大规模模拟的范式转变,为在原子尺度上理性设计高性能金属材料提供了关键的计算工具。

除上述通用设计平台外,还存在专门为传统金属材料(钢、铝、钛、镍基合金等)设计的平台。DAMASK[11]是一个统一的多物理场晶体塑性仿真工具,在晶体塑性的基础上,采用多种本构模型和均质化方法,利用连续介质力学边值问题求解材料各点的变形和应力的本构响应。此外,在中国材料基因组研究与集成计算平台的发展进程中,涌现出一系列具有代表性的软件与平台。中国科学院计算机网络信息中心主导开发的 MatCloud+材料云平台[12],提供了一个集成的在线计算环境,支持高通量计算、数据管理与分析的工作流自动化,极大提升了材料计算的效率与可重复性。这些高通量计算平台有望在探索金属材料的性能-结构构效关系研究中发挥关键作用,为新一代高性能金属材料的理性设计提供强大支撑。北京应用物理与计算数学研究所自主研制了应用于金属、合金、化合物等材料的软件程序,用于计算组分—结构—物性,并集成构建了 ProMe 极端条件下物性集成计算平台[13],具体工作包括:开发了基于对称性采样与机器学习的化学有序结构搜索程序 ABC,以及基于广义相似度函数的化学无序结构建模程序 SAE;提出了改进的平均场势方法并研制了高温自由能程序 MFP2,进而发展了高压热弹性方法并开发了高温弹性常数程序 HTEM;推导了相图解析模型并发展了耦合第一性原理计算、分子动力学和相图计算(calculation of phase diagram, CALPHAD)方法的多组元多相热力学建模方法,研制了相图自动建模软件 Auto-CALPHAD。由燕山大学亚稳材料全国重点实验室极端变形研究中心材料微结构与力学性能模拟研究室开发的极端条件下材料力学性能计算软件力牧(limo)[14],可用于单晶和多晶体系的建模与计算,它基于第一性原理计算参数,通过结合位错理论和过渡态理论,构建材料微结构到宏观力学性能的桥梁,实现了材料力学性能的计算。此外,还有可以动态模拟合金凝固、相变、再结晶等组织演化过程的相场计算软件 Powerbyte[15],多效应仿真 CAE 软件 GalaxyEDS 等。这些软件平台和方法成功应用于铝合金、铜合金、新型航空航天发动机高温合金以及高熵合金等材料的研发[13-14],显著提升了设计效率并缩短了研发周期。

1.2 材料数据库

在材料计算模拟能力飞速发展的同时,与之配套的材料数据库建设也取得了显著进展,它们共同构成了材料基因工程的核心基础设施。这些数据库通过系统收集、整理和管理海量的计算与实验数据,为材料智能设计提供了不可或缺的数据基础。随着材料研究进入“数据驱动”的新范式,材料科学数据库不再仅仅是存储数据的仓库,而是演变为支撑材料发现、设计和优化的智能中枢。这些数据库通过深度融合计算、实验与机器学习方法,正在从根本上改变探索和认知材料的方式。美国劳伦斯伯克利国家实验室和麻省理工学院等单位联合开发的 Materials Project[23]存储了几十万条第一性原理计算数据,体系包含金属、合金及化合物等,数据库中包含了大量缺陷和力

学性质,如形成能、表面能、晶界能、弹性系数、模量和泊松比等,为结构金属强化和脆化性能构效关系的研究提供了大量的精准数据。除此之外,杜克大学开发的 AFLOW[8]也存储了大量二元和多元合金的结构和性能数据,并提供了基于密度泛函理论、信息学数据挖掘和筛选策略。美国西北大学开发的 Open Quantum Materials Database(OQMD)[24]不仅包含大量的晶体结构、能带和能量等性质,还可利用元素及算法计算材料的相图,预测热力学稳定相。中国近年来在该领域也取得了长足的进步,涌现出一系列具有代表性的数据库平台。中国最大的材料基因工程数据库平台(Materials Genome Engineering Databases, MGED)[25]涵盖结构材料、能源材料、催化材料、特种合金、稀土功能材料、生物医用材料、核材料、复合材料、超导材料等;除数据库外,还拥有第一性原理计算平台、原子势函数库、数据挖掘系统、实验数据处理软件包、材料设计工具等多项功能。国家材料科学数据共享网[26]由国内 30 余家科研单位参与建设,以钢铁材料和先进合金材料为主,材料学科领域基础科学数据库[27]及 Matdat 等数据库中都包含了大量金属材料性质与性能。

2 描述符与理论模型

获得了高通量计算与数据库提供的海量原子结构、电子结构及力学性能数据后,研究的关键便从数据积累转向了知识挖掘。在此过程中,描述符与理论模型的构建起到了承上启下的核心作用,如图 3 所示。基于理论计算与材料数据库,结合传统的材料热动力学模型(如 Hall−Petch 关系、晶界偏析、晶界脆化模型等),构建电子、原子和几何描述符。在机理阐释层面,描述符(如原子局域环境特征、电子结构参数、晶格畸变等)作为量化材料微观特征的数学表达,是将宏观性能与微观结构直接关联的桥梁,用于揭示材料成分—结构—性能间的内在规律,形成清晰的物理图像;在性能预测层面,这些具有物理图像的描述符可直接预测材料性能,或作为机器学习模型的优质特征输入,显著提升了预测模型的准确性与外推能力,而理论模型则为数据驱动发现提供了物理约束与验证基准,共同构建了既保持物理正确性又具备高效预测能力的计算分析体系。这种“物理机理与数据科学”深度融合的研究模式,有效衔接了理论计算与工程应用,不仅深化了对材料本质的理解,更大幅加速了面向特定性能需求的新材料设计进程,是实现从“海量试错”到“理性设计”这一范式变革的理论核心。

以晶界性质及其相应的力学性能为例,传统模型多以晶界能、晶界偏析能、剪切模量等热力学性质作为关键描述符,桥接微观结构与宏观力学性能。例如,Rice−Wang 理论通过晶界偏析能和表面偏析能之差,建立了杂质元素偏析与晶界结合能之间的定量关

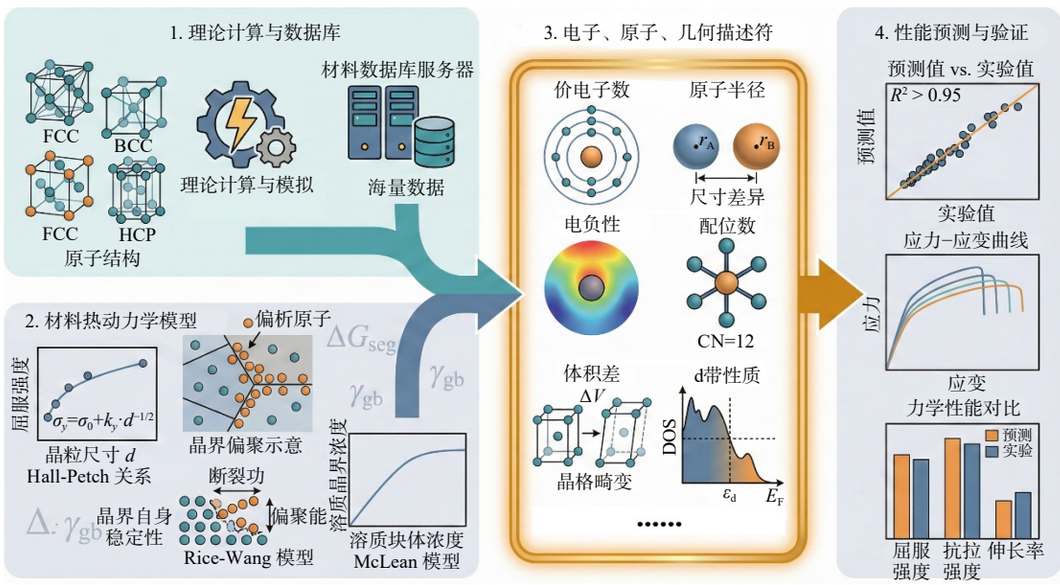


图 3 理论模型与描述符在晶界性质与材料性能预测中的应用

系,为理解晶界脆化提供了理论基础[28]。通过引入晶界能、晶界滑移速率、晶界体积分数、层错能等关键描述符,不仅能够基于经典的位错塞积理论和加工硬化模型构建 Hall−Petch 关系以预测常规晶粒尺度下的材料强度,而且能够有效描述晶界主导的变形机制,从而实现晶界强化行为的统一描述和准确预测[29]。西安交通大学刘刚和孙军团队以氢陷阱理论为指导,将界面氢结合能确立为关键描述符,成功设计并制备出具有超高氢结合能的核壳结构复合相,该界面作为强氢陷阱,有效抑制了氢在晶界等敏感位置的偏聚,从而在力学性能上同时实现了强度提升与氢脆敏感性大幅降低[30]。因此,通过选取具有明确物理意义的微观特征作为描述符,能够构建起连接不同尺度的理论模型,从而系统揭示晶界结构对金属力学性能及损伤行为的影响机制。

尽管传统模型通过晶界能、偏析能和剪切模量等热力学描述符,挖掘微观结构与宏观性能的物理图像,但这一范式存在明显的局限性。首先,从应用层面看,这些描述符本身(如晶界能、偏析能)通常是第一性原理计算或复杂实验测量的“输出结果”,而非易于获取的“输入参数”。这种特性导致它们难以直接用于快速、低成本的材料筛选与性能预测,即为了预测性能,需要先进行复杂的计算或测量来获得描述符,这极大地限制了其在实际工程设计中的应用。其次,从理论层面看,这类描述符的通用性不足,在简单的纯金属或二元体系中尚可计算,但当面对现代工程广泛应用的多组元复杂合金时,由于组元间复杂的交互作用,精确计算任意成分下的晶界偏析能和界面性质变得异常困难,使得传统模型难以有效指导多元合金的设计。因此,发展更易获取且具备良好扩展性的新描述符(即本征描述符),成为突破当前研究瓶颈的关键。

以晶界偏析及由偏析引起的晶界脆化为例,传统的晶界偏析模型如 McLean 模型假设偏析驱动力完全由弹性应变能 E_{elastic} 决定,可由母体材料的块体模量 K_{M}、溶质元素的剪切模量 G_{S} 以及母体原子和溶质原子的半径 r_{M} 和 r_{S} 描述,晶界偏析能 E_{seg} 计算公式如下:

$$E_{\mathrm{seg}}=E_{\mathrm{elastic}}=\frac{24\pi K_{\mathrm{M}}G_{\mathrm{S}}r_{\mathrm{S}}r_{\mathrm{M}}\left(r_{\mathrm{M}}-r_{\mathrm{S}}\right)^{2}}{3K_{\mathrm{M}}r_{\mathrm{M}}+4G_{\mathrm{S}}r_{\mathrm{S}}} \tag{1}$$

与之相似的热力学或半经验模型还有 Miedema 模型等,这些模型大多由弹性相互作用、溶解焓、纯金属界面能差异、弹性尺寸失配能等参数估算晶界偏析焓,不仅依赖较多的能量参数,且适用性和精度都有限[31]。在经典热力学理论框架下,研究者们提出了大量用于量化晶界偏析行为的理论模型。基于晶界偏析中经典的弹性相互作用和化学相互作用,Huber 从溶质原子和晶界位点体积差 $\Delta V_{\mathrm{GB},i}$ 来量化偏析能的弹性部分,考虑了偏析位点的配位数 ΔN_i 和键能 E_{bond},量化了偏析能的化学效应 $E_{\mathrm{seg,bond}}$,提出了 Voronoi 体积差和配位数差作为偏析能描述符,并提出了描述偏析能 E_{seg} 的解析公式:

$$E_{\mathrm{seg}}=E_{\mathrm{elastic}}+E_{\mathrm{seg,\,bond}}=-B\frac{V_X-V_M}{V_{\mathrm{M}}}\Delta V_{\mathrm{GB},i}+E_{\mathrm{bond}}\Delta N_i \tag{2}$$

式中,B 为溶质元素的体积模量,V_X 为溶质元素的体积,V_{M} 为母体的原子体积,该公式被证明适用于 Al 和 Mg 晶界中不同溶质[32−33]。

除 FCC(面心立方结构)和 HCP(密排六方结构)金属外,在 BCC(体心立方结构)金属中,反映双峰模态效应 ΔE_{d} 与 sp−d 轨道杂化效应 ΔE_{sp} 的电子结构描述符可用于构建晶界偏析能的量化模型[34],表示为

$$E_{\mathrm{seg}}\approx\Delta E_{\mathrm{d}}+\Delta E_{\mathrm{sp}}=a_1\Delta\mathrm{dip}+a_2x_{\mathrm{sp}} \tag{3}$$

式中,Δdip 为缺陷附近原子与远离缺陷的块体原子的局部态密度的双峰模态效应之差;x_{sp} 描述了 sp−d 杂化效应中缺陷部位的局部环境,可由 Voronoi 体积和 sp 带中心计算;对于每一种基体材料−溶质元素对,可以通过最小二乘法拟合系数 a_1 和 a_2,或局部线性回归等方法构建机器学习模型。

上述基于电子结构描述符的模型具有良好的有效性,但需要复杂的电子结构计算,面对成分更复杂的合金体系或高通量筛选时,需要更高效或易于获取的参数。

针对 Mo 基体系的研究表明,可利用金属原子半径与溶质/基体内聚能作为关键参数,构建能够有效预测晶界强化/脆化能 ΔE_{B} 的双因素线性模型[31],可表示为

$$\Delta E_{\mathrm{B}}=k_{\mathrm{coh}}E_{\mathrm{coh}}^{X}+k_{\mathrm{R}}\frac{R_X-R_{\mathrm{Mo}}}{R_{\mathrm{X}}} \tag{4}$$

式中,E_{coh}^{X} 为溶质元素的内聚能;R_X 为溶质元素半径;R_{Mo} 为 Mo 的原子半径;$\frac{R_X-R_{\mathrm{Mo}}}{R_X}$ 为相对原子半径,衡

量了溶质 X 的应力效应;k_R、k_{coh} 为由最小二乘法拟合的系数。

除此之外,本研究组的前期工作发现,溶质的价电子数 S_v、电负性 χ 和尺寸 R,以及晶界对应切割表面的广义配位数控制溶质在双晶晶界处的溶质偏析能[35],构建了溶质偏析的电子描述符 $D_{seg}=\left[\left(\frac{\sigma}{R}\right)^{12}-\left(\frac{\sigma}{R}\right)^{6}\right]\times\frac{S_v^{\,2}}{\chi}$ 和几何描述符 $\overline{CN}_{cs}$;针对多晶晶界偏析,通过结合应变和配位数概念的应变配位数 CN_{str} 可定量描述多晶晶界的溶质偏析能,证明了原子结构的无序化使得多晶晶界区和非晶体中金属间键合性质由金属键向离子键转变,结合溶质和基体的原子半径 r_s 和 r_m 以及基体的内聚能 $E_{coh,m}$,构建了量化多晶晶界溶质偏析的理论预测模型[36],具体表示为

$$E_{seg}=kCN_{str}+b=\left[\frac{1}{12}E_{coh,m}(30-CN)\left(\frac{r_s}{r_m}-1\right)+10\right]\cdot(CN_{str}-CN)+\theta \quad (5)$$

式中,k、b 为拟合系数,θ 为常数项。

从传统的热力学模型到近年发展的本征描述符方法,呈现出从宏观参数向微观机制深化的发展趋势,这些新型描述符基于热力学模型、原子间键合规律、电子结构性质等微观特征,建立了基于原子和电子结构的物理图像,揭示了偏析过程中弹性相互作用、化学相互作用、各向异性以及耦合规律等微观机制。这些基于微观机制建立的描述符不仅在不同金属体系中展现出良好的预测精度,更重要的是为理解晶界偏析机制提供了更深刻的物理见解,为实现晶界偏析行为和晶界脆化的精准预测提供了有效工具。

本征描述符与相应模型不仅在纯金属和简单二元合金中表现出良好的适用性,更在多组元合金体系,特别是高熵合金等复杂材料中展现出独特价值。这类基于电子与原子本征特性的描述符能够有效跨越巨大的成分空间,通过建立成分−微观结构−性能之间的定量关联,为在广阔化学空间中筛选具有优异力学性能的新型合金提供理论指导与高效设计工具。Johnson 等[37]基于经典的熔化热模型探讨材料的宏观性质与键能的关系,提出了间隙电子密度线性描述符 ρ_0,由电荷 Q_0、原胞体积 Ω_{cell} 和原子体积 Ω_{IS} 计算得到 $\rho_0=Q_0/(\Omega_{cell}-\Omega_{IS})$,可用于量化金属的模量和极限强度,通过简单的以浓度为权重的混合规则,构建了合金体系的间隙电子密度,用于预测高熵合金和非晶的极限强度与模量。本研究组的前期工作基于经典的紧束缚理论和断键模型,提出了基于内聚能 E_{coh} 和原子半径 r 的描述符来量化金属的键强,其中 E_{coh}/r^5 用于过渡金属等强键金属,$E_{coh}/r^{2.5}$ 用于主族金属等弱键金属,建立了晶界稳定性和力学性能的预测模型,适用于纯金属,还可扩展至高熵合金和非晶合金多种体系。不仅快速预测晶界能,还可量化模量、熔化热、玻璃化转变温度、极限强度和断裂强度等宏观性能,揭示了基于键强的物理图像[1]。针对 Hall−Petch 关系系数,结合反映原子键合性质的紧束缚理论和经典 Hall−Petch 关系的加工硬化模型,提出基于价电子数 S_v、电负性 χ、周期 N_p、族数 N_g 和内聚能 E_{coh} 的本征描述符,其中电子描述符 $\mathcal{K}=\left(\frac{N_p}{4}\right)^{(\sqrt{N_g}-3)}\times\frac{S_v^2}{\chi}$ 量化了 Hall−Petch 关系的晶格摩擦项,内聚能的平方 $E_{coh}^{\,2}$ 量化了 Hall−Petch 关系的晶界阻力系数,结合平均场效应还可扩展至合金体系,不仅具有较高的预测精度,还揭示了 Hall−Petch 系数在电子结构层面起源于 d 带态密度的一阶和二阶矩[38]。

基于电子结构与原子结构本征特性的描述符和理论模型研究为材料性能预测建立了重要的理论基础,构建了从微观特性到宏观性能的定量关联框架。这些具有物理意义的描述符不仅深化了对材料强化与脆化机理的理解,更重要的是为后续机器学习模型提供了可靠的特征输入,确保了数据驱动模型的可解释性和预测准确性,为最终建立可解释、可预测的构效关系奠定了坚实的理论基础,实现对多元合金界面性质及力学性能的高效预测。

3 机器学习方法与大模型

随着材料体系的日益复杂和对性能要求的不断提升,传统研究方法在处理多变量、非线性问题时面临严峻挑战。在这一背景下,机器学习方法凭借其强大的数据挖掘和模式识别能力,为材料科学研究开辟了全新的研究范式。如图 4 所示,前期的机器学习与人工智能应用以传统机器学习模型(常被称为“小模型”,如随机森林、支持向量机、梯度提升树等)为代表,基于大量的材料数据,采用多维描述符和多种机

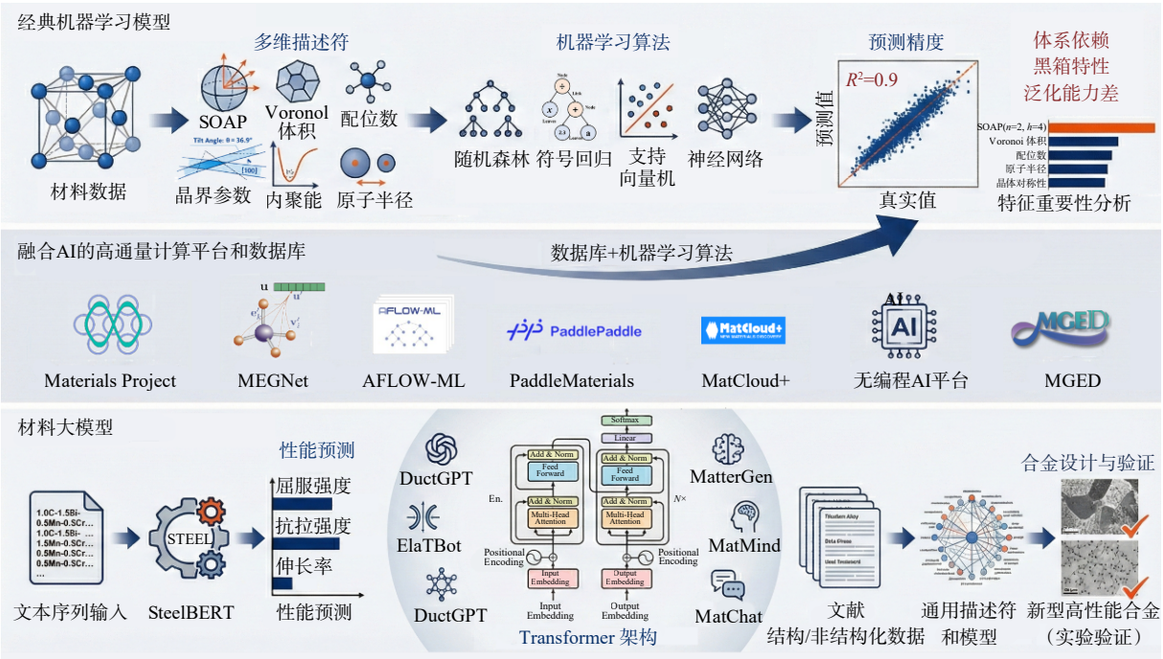


图 4　机器学习小模型与大模型示意

器学习算法，构建可快速预测材料性能的经典机器学习模型(即小模型)。这类模型善于处理明确的数值特征，在成分优化、性能预测等特定任务中表现出高效与精准。值得注意的是，小模型的效能高度依赖于高质量、标准化的数据基础。当前，机器学习工作流已深度集成于计算平台或数据库之中，实现了从“数据生成(高通量计算)→特征工程(描述符提取)→模型训练→性能预测”的自动化闭环，显著提升了研发效率。然而，其决策过程通常被视为“黑箱”，即其内部决策逻辑不透明，难以提供基于物理机制的解释。为突破这一局限，研究正朝着“白箱”(或“灰箱”)化方向发展。“白箱”模型核心在于实现模型的可解释性与物理一致性。具体体现为 2 个相互关联的方向：一是通过可解释机器学习方法(如符号回归等)揭示“黑箱”的决策依据，使预测结果与材料物理、化学机制相关联；二是推进“物理信息机器学习”，其关键在于物理约束内嵌——将原子键合性质、热力学与动力学等基本物理规律以数学形式嵌入模型架构或损失函数，从而约束输出自洽于科学原理。与此同时，基于 Transformer 架构的大语言模型(大模型)在材料领域迅速兴起。这类模型通过对海量学术文献、数据库等多源材料文本进行预训练，获得理解与生成材料专业语言的能力，服务于知识问答、文献挖掘和生成式设计等更复杂的认知任务，还与计算平台和大数据库深度集成，极大地拓展了人机协作的智能边界，显著提升了科研工作的交互效率与智能化水平。因此，材料数智化进程中的机器学习与人工智能的发展，呈现出从小模型的“黑箱”预测到可解释方法与基于物理信息的模型，再到大模型的发展方向。这些技术共同构建起一个可解释、可验证、可协作的材料智能研发体系，推动材料研究从经验试错加速迈向理性设计与自主发现。

3.1　机器学习小模型

以晶界偏析和强/脆化研究为例，传统机器学习模型展现出强大的预测能力。通过整合 Voronoi 参数、笛卡尔坐标、配位数、内聚能、价电子特性、原子受力和晶界结构参数等多维特征，可构建随机森林、极限梯度提升树和人工神经网络等机器学习模型，这些模型在 Mg 和 Al 等合金体系的 30 余种重合点阵模型晶界中表现出优异的预测性能[33,39]。结合原子位置平滑重叠(smooth overlap of atomic positions, SOAP)描述符与主成分分析的机器学习方法，可准确预测包含 15 种不同溶质–母体对的多晶体系偏析行为[40]。此外，基于内聚能、溶质–母体升华能、表面能及原子半径等关键参数构建的支持向量机模型，实现了对晶界脆化能的定量预测，为理解界面脆化机制提供了新途

径[41]。这些研究结果均表明,机器学习模型在预测晶界偏析和脆化行为方面展现出显著优势,但其预测精度严重依赖于输入描述符的合理选择与特征工程的优化设计。此外,机器学习模型广泛应用于合金甚至高熵合金缺陷和力学性能构效关系的挖掘,例如,Wen 等[42]结合传统的固溶强化模型,筛选了用于描述高熵合金固溶强化的因素如电负性、剪切模量、尺寸错配等。图卷积神经网络可用于量化金属的晶体结构,用于预测形成能、表面能、模量和弹性常数等性质[43]。目前,机器学习模型成功应用于相结构预测、显微组织预测、成分−工艺−性能预测和服役行为预测等。西安交通大学丁向东和孙军团队将位错动力学与断裂理论作为物理约束,嵌入到无监督机器学习框架的损失函数中,该算法能够从复杂的声发射信号中,精准分离出来自位错运动和微裂纹扩展等不同界面损伤机制的特征,从而实现对材料力学性能退化(如断裂)的动态早期预警,建立了从微观界面损伤到宏观失效的智能监测桥梁[44]。这些创新性的工作共同展示了现代材料研究的核心范式:以物理理论模型为指导,利用机器学习作为加速设计和高维信息解析的强大工具,最终通过对材料性质的精准调控或解析,实现材料宏观力学性能的突破或对其失效过程的前瞻性预测。

尽管这些模型采用了具有物理意义的描述符作为输入特征,其固有的“黑箱”特性仍导致模型难以提供对偏析和脆化机制的直观物理解释。这种机理阐释能力的欠缺,在一定程度上限制了机器学习方法在新材料体系探索中的理论指导价值,因此,可解释机器学习方法所代表的“白箱”模型,因其能提供直观的物理机制洞察,在材料科学研究中具备重要价值。例如,Jiang 等[45]通过符号回归方法,利用价电子距离 S、内聚能 E_{coh}、线性热膨胀系数 l_{t}、晶界能 γ_{GB} 和杨氏模量 E 构建了定量的 Hall−Petch 关系,直观地揭示了 Hall−Petch 系数的决定性因素,Hall−Petch 关系可表示为

$$\sigma=\frac{79E_{\mathrm{coh}}}{S^3\sqrt{l_{\mathrm{t}}}}+1.2\sqrt{\frac{\gamma_{\mathrm{GB}}E}{l_{\mathrm{t}}d}} \quad (6)$$

同样地,Wei 等[46]基于多任务 Lasso 算法和符号回归模型对钢的力学性能进行特征筛选,发现通过优化合金成分(提高 Cr、Mo、Ni、Mn 含量)与控制工艺参数(提高淬火温度、降低回火温度、减小夹杂物尺寸),可同步提升材料的疲劳强度、拉伸强度、断裂强度及硬度等综合力学性能。符号回归等可解释机器学习方法,通过构建具有物理意义的数学表达式,不仅实现了准确的性能预测,更重要的是揭示了影响材料性能的关键因素及其内在作用机制。这类“白箱”模型为理解材料构效关系提供了重要支撑,显著提升了机器学习在材料设计中的理论指导价值。

3.2 高通量计算平台集成机器学习模块

机器学习模型通常需要训练庞大的数据集,因此与高通量计算平台的深度集成变得至关重要。例如,Materials Project 平台在提供丰富材料数据库的基础上,其集成的机器学习模块已广泛应用于金属界面性质和力学性能的预测研究中[23]。MEGNet(Materials Graph Network)框架直接将晶体结构或分子结构转换为图数据,无需复杂的人工特征工程,提供在庞大的 Materials Project 数据集上预训练好的模型,用户可直接用于预测新材料的形成能等关键性质[47]。上海大学研究团队开发了无需编程的人工智能平台,采用图形化用户界面,集成了数据预处理、特征工程、模型训练与性能评估等完整工作流,支持随机森林、支持向量机、神经网络等多种经典算法,可应用于高性能钢材料、高熵合金等体系的筛选[48]。PaddleMaterials 作为百度飞桨团队推出的材料科学计算套件,采用多层级模块化架构,集成了晶体图卷积神经网络(crystal graph convolutional neural network, CGCNN)、MEGNet、晶体哈密顿图神经网络(crystal Hamiltonian graph neural network, CHGNet)等先进模型,支持晶体结构预测、分子动力学模拟和电子密度计算等核心功能[49],该平台通过物理机理驱动、数据驱动及数理融合 3 种求解模式,在二维材料构建、核磁共振(nuclear magnetic resonance, NMR)谱图解析等实际应用中展现出接近密度泛函理论(density functional theory, DFT)精度的计算能力与显著效率优势,为材料研发提供了从属性预测到结构生成的端到端解决方案。第 1 章节中提及的高通量计算平台如 MatCloud+材料云平台和中国最大的材料基因工程数据库平台 MGED 等都涵盖了机器学习处理模块。

3.3 材料大语言模型

随着机器学习在材料科学研究中的深入应用,传统的“小模型”(如针对特定性能的回归模型)虽然在

特定数据集上表现出良好的预测性能,但其局限性也日益凸显:模型泛化能力受限、可迁移性差,且难以捕捉材料体系中复杂的高阶非线性关系。这些模型通常在特定合金体系或有限晶界类型上训练得到,当其应用于具有不同成分或晶体结构的其他材料体系时,很难具备较好的预测能力。例如,利用 SOAP 描述符的机器学习模型只能应用于一种溶质–母体对,换句话说,对于不同的溶质–母体对需要训练对应的数据得到不同的机器学习模型。这种体系依赖特性主要源于 2 个方面:一是不同材料体系的原子间相互作用存在本质差异,缺乏量化合金化效应的描述符,导致现有描述符与偏析能之间的映射关系发生变化;二是模型训练所依赖的数据集往往难以覆盖所有可能的材料空间。这一局限性严重制约了机器学习模型在多元合金设计中的普适性应用,亟需发展具有跨体系预测能力的新型机器学习框架。在此背景下,材料科学领域正迎来向“大模型”范式转变的重要时刻。大语言模型是基于 Transformer 架构的深度学习模型,其核心在于通过自注意力机制处理序列数据,能够从海量材料数据中学习复杂的构效关系。材料科学领域的大模型基于商业大模型发展而来,利用自注意力与多头注意力机制,能有效捕捉材料数据中元素(如原子、官能团)间的复杂依赖关系和信息。通过前馈层的非线性变换和位置编码对序列/空间信息的嵌入,模型能够理解与表征材料科学的文献/语料库,将非结构化的文本化工艺路线和化学成分转换为数学向量,构建模型并预测合金的力学性能[50−51]。自回归学习的 Transformer 架构使模型能够根据现有信息逐步生成新的、合理的数据序列,这直接应用于生成新材料结构、预测合成步骤以及模拟动态过程,为材料设计与发现提供了强大的生成和预测工具。在模型设计阶段,需以 Transformer 为骨架,针对材料数据特点引入领域知识,如将物理化学约束融入损失函数;最后通过预训练–微调范式,先使用无标注材料数据进行学习,再在特定任务数据上进行有监督训练,期间采用分布式计算和混合精度等技术优化训练效率。这种专业化的训练方法确保了材料大模型既保持了 Transformer 架构的强大表征能力,又满足了材料科学研究对准确性、可解释性和物理一致性的严格要求,为加速新材料发现提供了有效的智能工具[51]。材料大模型通过在海量跨尺度、多类型材料数据上进行预训练,能够学习到更普适的材料特征表示,从而实现对不同材料体系、多种性能指标的协同预测与跨任务迁移。这种基于大模型的新范式,不仅有望突破传统机器学习模型在应用范围上的瓶颈,更为构建覆盖“成分—结构—工艺—性能”全链条的材料智能设计平台奠定了坚实基础。

首先,大模型可基于领域内文献文本预训练构建预测模型。北京科技大学开发的 SteelBERT 大模型,创新性地将大模型应用于钢材设计领域,提出了一套从材料文本到钢材性能的端到端流程。预训练语料包含 420 万材料科学摘要和 5.5 万篇钢相关文章,通过将钢材成分和制造工艺文本序列作为输入,SteelBERT 将化学成分与工艺文本转换为 768 维向量。该模型可准确预测屈服强度、抗拉强度和伸长率等关键力学性能。在针对奥氏体不锈钢的特定应用中,仅使用 64 个样本进行微调后就实现了更高精度的预测,并成功指导工艺优化,获得细化的晶粒尺寸,实现了优异屈服强度、抗拉强度和伸长率等综合性能[52]。香港大学开发的 ElaTBot 大语言模型面向材料弹性常数张量预测和材料反向设计,该模型通过融合自然语言处理与材料科学知识,构建了能够同时完成弹性常数张量预测、体积模量计算和新材料生成的多功能智能体,ElaTBot 在弹性常数预测中的误差较同类模型降低 33.1%,并借助检索增强生成技术实现了小样本条件下的高精度推断[53]。

除基于文本与训练的模型外,大模型还可整合并嵌入物理信息,提升模型的可信度与科学合理性。美国艾姆斯国家实验室开发了 DuctGPT,通过融合物理信息与生成式 Transformer 架构,实现了难熔多主元合金韧性的高精度、高通量预测,引入基于密度泛函理论计算的物理描述符如电子态密度、d 轨道杂化等关键描述符作为输入,既保持了模型的预测精度,又确保了结果的物理可解释性,揭示了电子结构对本征韧性的主导作用[54]。这些工作均表明,将具有明确物理意义的描述符与大模型深度融合,是构建兼具高预测精度与强物理可解释性材料设计模型的有效范式,为实现材料性能的精准预测和高通量筛选开辟了新路径。这类物理信息驱动的生成式设计模型还有微软研究院开发的 MatterGen 模型[55]等,有望应用于材

料微观结构与力学性能构效关系的建立与预测。同时,这类物理信息内嵌的大语言模型的发展表明,其预测能力的科学根基与可靠性高度依赖于底层物理理论的支撑,描述符与理论模型也成为了这类大模型抵抗可能产生物理不一致、科学失准、信息不存在或无效预测的风险(即“科学幻觉”)、实现可解释设计与性能精准预测的核心。

大模型的预测可靠性根植于物理描述符与理论框架提供的约束与指导另一方面,其强大的数据融合与生成能力,又为发现新描述符、构建新理论提供了全新的路径。西北工业大学通过融合大型语言模型与材料领域知识,创新性地提出了一种融合非结构化数据与结构化数据、自动构建通用描述符的钛合金设计框架,该框架基于 Transformer 架构,通过对超过 4.6 万篇文献摘要进行预训练,成功提取出能够同时预测多种力学性能的丰富描述符。其核心创新在于,该模型能够将合金成分(如“Ti-6Al-4V”)自动映射为一个高维语义向量,即“增强描述符”。这一过程并非简单的数值编码,而是模型基于从海量文本中学到的合金化规律、相稳定性关联等知识,对成分进行的深度语义表征。研究进一步通过遗传算法,从 768 维的语义嵌入中筛选出最具预测力的 13 维核心描述符。实验表明,基于此类“增强描述符”构建的预测模型,在屈服强度、抗拉强度等关键性能指标上的回归系数较传统方法提升超过 30%,通过钼当量理论和多目标优化,从百万级候选合金中筛选出多个高性能钛合金配方,并经过实验验证其性能优越性[56]。上述这些模型均表明,理论模型/描述符与大模型之间的双向闭环特点,为材料科学的数智化研究提供了全新范式。

此外,大模型还可作为动态、交互式的智能体与对话系统,成为一个强大的跨领域知识查询与信息整合工具。例如上述提及的 ElaTBot 大语言模型,还构建了一个多功能的智能体系统(ElaTBot Agent),提供了基于 Gradio 的对话界面,用户无需编程即可进行材料性能预测与逆向设计。除此之外,还有上海硅酸盐研究所的 MatMind 材料科学设计的先进大模型[57]和松山湖材料实验室的 MatChat 面向材料研发的智能对话系统等[58]。这些智能对话系统能够理解用户以自然语言描述的材料设计目标、工艺条件或性能需求,并据此调用内部的模型与数据资源,提供从成分初步筛选、工艺参数建议到性能初步评估的整体流程。这些大模型工作为实现材料设计的自动化与智能化提供了创新范式,显著提升了新材料研发效率。

4 结论

在传统金属材料的缺陷性质与力学性能研究中,高通量计算、理论模型与描述符、机器学习与大模型三者构成了一个完整的闭环系统。高通量计算作为系统的起点,通过多尺度模拟方法,批量生成包含成分、结构和性能的数据,为整个闭环提供丰富的数据基础。在此基础上,通过物理分析和数学建模,提炼出基于材料本征性质的关键描述符,构建起连接微观机制与宏观性能的定量关系,为金属材料性质与行为提供理论解释;同时理论模型可为机器学习模型提供关键的物理约束,为黑箱式机器学习提供物理图像,促进机器学习模型的可移植性和通用性。机器学习与大模型能够基于高通量计算产生的数据和描述符,建立精准的预测模型,智能筛选最具价值的新计算样本,指导高通量计算的方向;同时利用特征工程,识别关键描述符,反馈优化理论模型。这一闭环实现了“数据生成→理论提炼→智能优化→精准指导”的完整循环。每个环节的输出都成为下一环节的输入,形成持续改进的螺旋式上升,推动材料设计从经验驱动向数据驱动的范式转变。本综述的目的在于系统性地提出并阐释了“高通量计算—理论模型与描述符—机器学习与大模型”3 者构成的闭环研究范式,并强调其在材料数智化研究中的协同性与不可或缺性。相较于以往较多关注单一技术(如机器学习预测或高通量筛选)的综述,本文着重论述了计算、理论与智能 3 者如何相互驱动、相互校验,共同构成从数据到知识再到设计的完整链条。

尽管材料数智化研究展现出巨大潜力,但其发展仍面临如何贯穿数据、理论与智能核心的深层挑战。数据来源分散、格式不一、质量参差使得研究所需的高精度、标准化材料数据集仍然稀缺,跨尺度计算和模拟仍是材料计算领域长期存在的核心难题。在理论模型方面,兼具高普适性与高精度的理论模型仍匮乏,难以在广泛适用性、定量预测精度及物理机制清

晰性之间取得理想平衡。机器学习模型普遍存在跨体系迁移能力不足的问题,难以构建微观组织/形貌与力学性能等复杂构效关系。而生成式大模型强烈依赖于高质量、特定领域数据,在复杂材料系统中缺乏物理可解释性,其能力还伴随着“科学幻觉”,即可能产生物理不一致、科学失准、信息不存在或无效预测的风险。

当前以数据驱动为核心的“计算—理论—机器学习”闭环,其未来演进的关键在于从“数据驱动”迈向“知识嵌入”的智能融合。未来的模型将不再是被动拟合数据的黑箱,而是内嵌物理定律与材料知识的“白箱”系统;通过在机器学习架构中显式引入解析描述符和模型、物理约束等,并利用可解释机器学习自动发现新物理规律,模型将同时具备高精度预测能力和科学发现潜力,解决模型的可靠性、可解释性与跨体系迁移性问题。在大模型层面,内嵌的物理规律将作为强约束,从生成机理上抑制“科学幻觉”。在数据层面,通过建立统一、标准的材料数据规范与交换协议,整合分散的数据库并实现跨平台互操作,有望改善数据质量与一致性。最终,通过“高质量标准化数据基础”与“强物理约束智能内核”的协同建设,推动材料数智化研究构建起更坚实、更可信的新范式。未来的闭环还将深度融合机器人实验、高通量制备与实时表征技术,形成一个“计算设计→机器人合成→自动表征→数据反馈”的完整加速环。理论模型与AI将不再仅输出一个静态的性能预测值,而是生成一套包含最优合成路径、关键工艺窗口,甚至预期显微组织形貌的、可执行的全流程实验方案。这标志着材料研究将从传统的“试错式实验验证理论”彻底转向“理论精准指导并驱动实验”,将新材料从概念到实证的周期从数年缩短至数月甚至数周,极大加速面向国家重大需求的材料攻关进程。

参考文献(References)

[1] Li X, Wu H, Gao W, et al. A roadmap from the bond strength to the grain−boundary energies and macro strength of metals[J]. Nature Communications, 2025, 16: 615.

[2] Zheng H, Li X G, Tran R, et al. Grain boundary properties of elemental metals[J]. Acta Materialia, 2020, 186: 40−49.

[3] Hirel P. Atomsk: A tool for manipulating and converting atomic data files[J]. Computer Physics Communications, 2015, 197: 212−219.

[4] Cheng J L, Luo J, Yang K S. Aimsgb: An algorithm and open−source Python library to generate periodic grain boundary structures[J]. Computational Materials Science, 2018, 155: 92−103.

[5] Hjorth L A, Jrgen M J, Blomqvist J, et al. The atomic simulation environment: A Python library for working with atoms[J]. Journal of Physics: Condensed Matter, 2017, 29(27): 273002.

[6] Ong S P, Richards W D, Jain A, et al. Python Materials Genomics (pymatgen): A robust, open−source Python library for materials analysis[J]. Computational Materials Science, 2013, 68: 314−319.

[7] Janssen J, Surendralal S, Lysogorskiy Y, et al. Pyiron: An integrated development environment for computational materials science[J]. Computational Materials Science, 2019, 163: 24−36.

[8] Curtarolo S, Setyawan W, Hart G L W, et al. AFLOW: An automatic framework for high−throughput materials discovery[J]. Computational Materials Science, 2012, 58: 218−226.

[9] Wang H, Zhang L F, Han J Q, et al. DeePMD−kit: A deep learning package for many−body potential energy representation and molecular dynamics[J]. Computer Physics Communications, 2018, 228: 178−184.

[10] Zhang Y Z, Wang H D, Chen W J, et al. DP−GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models[J]. Computer Physics Communications, 2020, 253: 107206.

[11] Roters F, Diehl M, Shanthraj P, et al. DAMASK–The Düsseldorf Advanced Material Simulation Kit for modeling multi−physics crystal plasticity, thermal, and damage phenomena from the single crystal up to the component scale[J]. Computational Materials Science, 2019, 158: 420−478.

[12] Yang X Y, Wang Z G, Zhao X S, et al. MatCloud: A high−throughput computational infrastructure for integrated management of materials simulation, data and resources[J]. Computational Materials Science, 2018, 146: 319−333.

[13] Gao X Y, Wang W Y, Chen X, et al. ProME: An integrated computational platform for material properties at extremes and its application in multicomponent alloy design[J]. Materials Genome Engineering Advances, 2025, 3(3): e70029.

[14] Lei M Y, Sun G P, Yang G C, et al. A computational mechanical constitutive modeling method based on thermally−activated microstructural evolution and strengthening mechanisms[J]. International Journal of Plasticity, 2024, 173: 103881.

[15] PowerByte [EB/OL]. (2024−01−15) [2026−09−15]. https://powerbyte.online/index/main/.

[16] Liu X H, Chen M H, Li P F, et al. Introduction to first−principles simulation package ABACUS based on systematically improvable atomic orbitals[J]. Acta Physica Sinica, 2015, 64(18): 187104.

[17] Xu K, Bu H K, Pan S N, et al. GPUMD 4.0: A high−performance molecular dynamics package for versatile materials simulations with machine−learned potentials[J]. Materials

Genome Engineering Advances, 2025, 3(3): e70028.
[18] Li X G, Chen C, Zheng H, et al. Complex strengthening mechanisms in the NbMoTaW multi−principal element alloy[J]. npj Computational Materials, 2020, 6: 70.
[19] Chang Z, Feng L, Xue H T, et al. Deep−learning potential molecular dynamics study on nanopolycrystalline Al–Er alloys: Effects of Er concentration, grain boundary segregation, and grain size on plastic deformation[J]. Journal of Chemical Information and Modeling, 2025, 65(7): 3282−3293.
[20] Khazieva E O, Shchelkatchev N M, Tipeev A O, et al. Accuracy, performance, and transferability of interparticle potentials for Al–Cu alloys: Comparison of embedded atom and deep machine learning models[J]. Journal of Experimental and Theoretical Physics, 2023, 137(6): 864−877.
[21] Suárez−Recio J, Piaggi P M, Domínguez−Gutiérrez F J, et al. Early stages of self−healing at tungsten grain boundaries from *ab initio* machine learning simulations[J]. Communications Materials, 2025, 6: 236.
[22] Sun H J, Zhang C, Tang L, et al. Molecular dynamics simulation of Fe−Si alloys using a neural network machine learning potential[J]. Physical Review B, 2023, 107(22): 224301.
[23] Jain A, Ong S P, Hautier G, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation[J]. APL Materials, 2013, 1: 011002.
[24] Kirklin S, Saal J E, Meredig B, et al. The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies[J]. npj Computational Materials, 2015, 1: 15010.
[25] MGED [EB/OL]. (2018−05−14) [2026−09−15]. http://www.mgedata.cn.
[26] 国家材料数据共享网[EB/OL]. (2020−01−02) [2026−09−15]. http://www.materdata.cn.
[27] 材料科学基础数据库[EB/OL]. (2021−05−28) [2026−09−15]. http://www.matsci.csdb.cn.
[28] Rice J R, Wang J S. Embrittlement of interfaces by solute segregation[J]. Materials Science and Engineering: A, 1989, 107: 23−40.
[29] Figueiredo R B, Kawasaki M, Langdon T G. Seventy years of Hall−Petch, ninety years of superplasticity and a generalized approach to the effect of grain size on flow stress[J]. Progress in Materials Science, 2023, 137: 101131.
[30] Jiang S Y, Xu Y T, Wang R H, et al. Structurally complex phase engineering enables hydrogen−tolerant Al alloys[J]. Nature, 2025, 641(8062): 358−364.
[31] Tran R, Xu Z H, Zhou N X, et al. Computational study of metallic dopant segregation and embrittlement at molybdenum grain boundaries[J]. Acta Materialia, 2016, 117: 91−99.
[32] Huber L, Grabowski B, Militzer M, et al. Ab initio modelling of solute segregation energies to a general grain boundary[J]. Acta Materialia, 2017, 132: 138−148.
[33] Huber L, Hadian R, Grabowski B, et al. A machine learning approach to model solute grain boundary segregation[J]. npj Computational Materials, 2018, 4: 64.
[34] Hu Y J, Zhao G, Zhang B Y, et al. Local electronic descriptors for solute−defect interactions in bcc refractory metals[J]. Nature Communications, 2019, 10: 4484.
[35] Li X, Liu P T, Gao W, et al. A universal descriptor to determine the effect of solutes in segregation at grain boundaries[J]. Journal of Materials Science & Technology, 2025, 222: 22−27.
[36] Wu H, Li X, Gao W, et al. Unconventional solute−matrix bonding at metallic polycrystal grain boundaries and their amorphous counterparts[J]. Acta Materialia, 2025, 301: 121590.
[37] Johnson D D, Singh P, Smirnov A V, et al. Universal maximum strength of solid metals and alloys[J]. Physical Review Letters, 2023, 130(16): 166101.
[38] Li X, Gao W, Jiang Q. Understanding the intrinsic framework of the Hall−Petch relationship of metals from the view of the electronic−structure level[J]. Acta Materialia, 2025, 292: 121071.
[39] Mahmood Y, Alghalayini M, Martinez E, et al. Atomistic and machine learning studies of solute segregation in metastable grain boundaries[J]. Scientific Reports, 2022, 12: 6673.
[40] Wagih M, Larsen P M, Schuh C A. Learning grain boundary segregation energy spectra in polycrystals[J]. Nature Communications, 2020, 11: 6376.
[41] Wu X B, Wang Y X, He K N, et al. Application of machine learning to predict grain boundary embrittlement in metals by combining bonding−breaking and atomic size effects[J]. Materials, 2020, 13(1): 179.
[42] Wen C, Wang C X, Zhang Y, et al. Modeling solid solution strengthening in high entropy alloys using machine learning[J]. Acta Materialia, 2021, 212: 116917.
[43] Xie T, Grossman J C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties[J]. Physical Review Letters, 2018, 120(14): 145301.
[44] Gou B Y, Chen Y, Xu S H, et al. Detecting deformation mechanisms of metals from acoustic emission signals through knowledge−driven unsupervised learning[J]. Nature Communications, 2025, 16: 6877.
[45] Jiang L, Fu H D, Zhang H T, et al. Physical mechanism interpretation of polycrystalline metals' yield strength *via* a data−driven method: A novel Hall–Petch relationship[J]. Acta Materialia, 2022, 231: 117868.
[46] Wei Q H, Xiong J, Sun S, et al. Multi−objective machine learning of four mechanical properties of steels[J]. Scientia Sinica Technologica, 2021, 51(6): 722−736.
[47] Treacher A H, Garg P, Davenport E, et al. MEGnet: Automatic ICA−based artifact removal for MEG using spatiotemporal convolutional neural networks[J]. NeuroImage, 2021, 241: 118402.
[48] Ma J X, Cao B, Dong S Y, et al. MLMD: A programming−free AI platform to predict and design materials[J]. npj Computational Materials, 2024, 10: 59.
[49] PaddleMaterials: A deep learning toolkit based on Paddle Paddle for material science [EB/OL]. (2025−08−13) [2026−

09−15]. https://github.com/PaddlePaddle/PaddleMaterials.
[50] Jiang X, Wang W R, Tian S H, et al. Applications of natural language processing and large language models in materials discovery[J]. npj Computational Materials, 2025, 11: 79.
[51] Wang G J, Hu J J, Zhou J, et al. Knowledge−guided large language model for material science[J]. Review of Materials Research, 2025, 1(2): 100007.
[52] Tian S H, Jiang X, Wang W R, et al. Steel design based on a large language model[J]. Acta Materialia, 2025, 285: 120663.
[53] Liu S Y, Wen T Q, Ye B L, et al. Large language models for material property predictions: Elastic constant tensor prediction and materials design[J]. Digital Discovery, 2025, 4(6): 1625−1638.
[54] Guduru S P R, Kekung M O, Ott R T, et al. DuctGPT: A generative transformer for forward screening of ductile refractory multi−principal element alloys[J]. Acta Materialia, 2026, 304: 121763.
[55] Zeni C, Pinsler R, Zügner D, et al. Mattergen: A generative model for inorganic materials design [J]. Arxiv preprint, 2023: arxiv: 2312.03687.
[56] Wang P, Jiang Y, Liao W J, et al. Generalizable descriptors for automatic titanium alloys design by learning from texts *via* large language model[J]. Acta Materialia, 2025, 296: 121275.
[57] MatMind [EB/OL]. (2025−04−07) [2026−09−15]. http://47.109.82.114:81/main.
[58] Chen Z Y, Xie F K, Wan M, et al. MatChat: A large language model and application service platform for materials science[J]. Chinese Physics B, 2023, 32(11): 118104.

Progress of the digital intelligence of metal grain boundary properties and corresponding mechanical properties

LI Xin, ZHANG Zhiwei, GAO Wang*, JIANG Qing

Key Laboratory of Automobile Materials, Ministry of Education, School of Materials Science and Engineering, Jilin University, Changchun 130022, China

Abstract Grain boundaries are one of the common defects in metallic materials and strongly affect the mechanical properties of metals. In response to the urgent demand for high strength and damage resistance of metallic materials in engineering applications, this paper takes the properties of grain boundaries and corresponding mechanical properties as examples to systematically review the digital and intelligent research progress of the microstructure and mechanical properties of metallic materials. This review elaborates on the closed−loop research paradigm of "high−throughput computing − theoretical model − machine learning and large language model". In detail, the high−throughput computing platform generates material data in batches through multi−scale computing, forming the large database. Based on the large database, combined with the classical thermodynamic/dynamic models of materials science and the bonding theories of metal physics, descriptors and theoretical models are constructed to establish the quantitative structure−activity relationship between microstructure and macroscopic mechanical properties. Machine learning can be used for rapid prediction of material properties. They can also integrate descriptors with theoretical models to solve the "black box" problem, thereby guiding theoretical calculations and experimental studies. Large−language models can be trained on massive materials science texts, demonstrating powerful cross−disciplinary knowledge integration and generation capabilities, and can serve as efficient intelligent research assistants to assist in material discovery and design. Finally, we also look forward to the development of this field towards the deep integration of physical constraint embedding and autonomous scientific discovery. We also propose the key challenges, such as data quality, theoretical universality, model transferability, and "Hallucination" of the large−language models. All these reviews provide a systematic theoretical reference for the digital and intelligent research on the grain boundary properties and corresponding mechanical properties of metallic materials, as well as for the design of high−performance metallic materials.

Keywords high−throughput computing; material databases; theoretical models; descriptors; machine learning; large models; metal grain boundary; mechanical properties

（责任编辑　魏来）