arXiv:2602.03069v1 [cs.DB] 3 Feb 2026

基于技能的自主智能体用于材料蠕变数据库构建

Yue Wu[a, *] 上海大学材料基因组研究院，中国上海 200444 repeatsongyue@gmail.com

Tianhao Su[a, *] 上海大学材料基因组研究院，中国上海 200444 thsu0407@gmail.com

Shunbo Hu[a, b, †] 上海大学材料基因组研究院[b] 上海大学文化遗产保护研究院、量子科学与技术研究所、硅酸盐质文物保护重点实验室（教育部），中国上海 200444 shunbohu@shu.edu.cn

Deng Pan[a, †] 上海大学材料基因组研究院，中国上海 200444 DPan_MGI@shu.edu.cn

摘要

数据驱动材料科学的发展目前受到一个根本性瓶颈的制约：绝大多数历史实验数据仍封存在传统科学文献中未结构化的文本和栅格化图表里。人工整理这些知识不仅耗费大量人力，而且容易出现人为错误。为应对这一挑战，我们提出了一种由大语言模型（LLM）驱动的自主智能体框架，旨在无需人工干预即可从科学 PDF 中挖掘高保真数据集。通过采用模块化的“技能型”架构，该智能体能够统筹复杂的认知任务，包括语义筛选、多模态信息提取和物理信息引导的验证。我们通过构建一个材料蠕变力学数据库，展示了该框架的有效性；这一领域具有复杂的图形轨迹和异质的本构模型。将该流水线应用于243篇出版物后，智能体在图形数据数字化方面实现了超过90%的验证后提取成功率。尤为重要的是，我们提出了一种跨模态验证协议，证明智能体能够自主将视觉提取的数据点与文本提取的本构参数（$R^2 > 0.99$）进行匹配，从而确保数据库在物理上的自洽性。这项工作不仅为研究不同材料体系中的时变变形提供了关键资源，也确立了一种可扩展的自主知识获取范式，为新一代自驱动实验室铺平了道路。

*这些作者对本研究贡献相同。

†通讯作者。

1 引言

在人工智能（AI）能力迅速发展的推动下，科学研究格局正经历深刻变革。“人工智能 for Science”（AI4S）已从理论上的可能性转变为实践范式，从根本上改变了假设的生成与验证方式[1, 2]。大语言模型（LLM）是这场变革的核心，它们不仅在自然语言处理方面表现出色，在逻辑推理和代码生成方面也展现出卓越能力[3, 4]。早期模型主要专注于文本生成，而近年来思维链（CoT）推理[5] 与 ReAct 范式[6] 的结合，已使大语言模型从被动生成器演变为自主智能体。这些智能体如今配备了模块化“技能”和工具使用能力，能够与外部环境交互[7, 8]。这一演进有望自动化以往仅由人类专家承担的复杂认知任务，并有可能以前所未有的速度加速科学发现[9, 10, 11]。

尽管技术取得了长足进步，人工智能在许多专业科学领域的应用仍面临一个持续存在的障碍：高质量数据[12]的匮乏。通用模型受益于互联网规模数据集，而材料科学、力学和化学等领域的领域特定任务，则依赖于往往零散、非结构化且隐藏在学术文献[13]文本和图表中的数据。在这些领域构建高保真数据集，一向既耗费大量人力，成本也高得令人难以承受。ChemDataExtractor[14]和 MatScholar[15] 等早期工作利用基于规则或基于 LSTM 的方法取得了显著进展，但在复杂语义推理和跨模态对齐方面仍常常力有不逮。人工整理依然是常态，却成本高昂，且容易出现人为错误[16]。这种“数据瓶颈”是阻碍数据驱动机器学习（ML）技术在这些关键领域广泛应用的主要障碍[17]。

然而，智能体框架和基于技能的架构的迅速成熟，为这一困境提供了新的解决方案[18]。值得注意的是，在专业领域获取相关文献绝非简单的检索任务；这一过程十分复杂，需要深厚的领域知识，才能从无关提及中辨别有效的实验情境，并评估各来源的方法学相关性。传统的自然语言处理（NLP）或光学字符识别（OCR）流水线缺乏跨模态推理能力，无法解决图形数据点与文本物理参数（例如温度 T、应力σ）之间的语义耦合问题[19, 20, 21, 22, 23, 24], ；而现代智能体能够模拟专家级认知。我们认为，这些框架目前已发展到足够成熟的水平，能够自主遨游浩瀚的科学文献海洋。通过定义特定技能，我们可以赋能智能体实现端到端自主性，在完全无需人工介入的情况下，自动收集相关论文、以专业判断筛选质量、提取各个数据点，并将这些信息整理成可查询数据库。为验证这一假设，我们聚焦于材料蠕变力学领域。选择这一领域，不仅是将其作为检验我们框架的试验场，也是因为它代表了一类典型的高维、非线性且稀缺的科学数据[25]。这一独特特性使其成为验证自主智能体物理信息推理能力的理想试金石，而非仅仅充当随机测试案例。NIMS 蠕变数据表[26]等权威存储库虽然提供了高保真基准数据，但从根本上依赖细致的人工整理和标准化测试规程。这种依赖使其本质上静态且范围有限，无法跟上学术文献中持续涌现的大量异质实验结果。因此，缺少一个同时包含蠕变曲线及其明确本构表达式的动态大规模数据库，严重制约了研究进展。一方面，从人工智能的角度看，这种匮乏阻碍了符号回归任务[27, 28] 和物理信息神经网络（PINNs）[29, 30]的发展；如果没有丰富的已知函数形式和实验数据，算法就难以发现或改进支配材料变形的物理定律。另一方面，对于实验研究人员而言，缺少集中式数据库使基准测试变得十分耗时，无法高效地将新的实验结果与历史数据进行比较。因此，建立这一领域的数据库，将有助于通过数据驱动的方式发现本构模型，并加快对高温材料性能的评估。

在此，我们展示了一种利用基于大语言模型的智能体与技能框架，自动化构建领域专用数据库的新型工作流程。以蠕变力学作为代表性案例研究，我们设计了一组专用技能，使智能体能够识别相关文献，并提取蠕变曲线及其明确的数学表达式。我们的工作阐明了大语言模型与基于技能的架构之间的协同作用如何突破数据瓶颈，为构建下一代 AI4Science 所需的基础数据集提供可扩展的路径。

2 方法论

自动化文献挖掘流程的总体架构如图 1 所示。该系统旨在将非结构化科学 PDF 转换为包含蠕变行为和本构模型的结构化、可查询数据库。该工作流程以基础大语言模型（大语言模型）作为中央推理智能体，统筹包含五个阶段的流程：（1）文献收集；（2）自动筛选；（3）多模态信息提取；（4）公式验证；（5）结构化存储。这种模块化设计确保文本元数据和图形实验数据均能以高保真度和物理一致性得到处理。具体而言，该框架的核心认知引擎基于 Qwen3-235B-A22B 基础模型[31], 构建，该模型提供了驱动这些复杂自主工作流程所必需的高级逻辑推理能力。

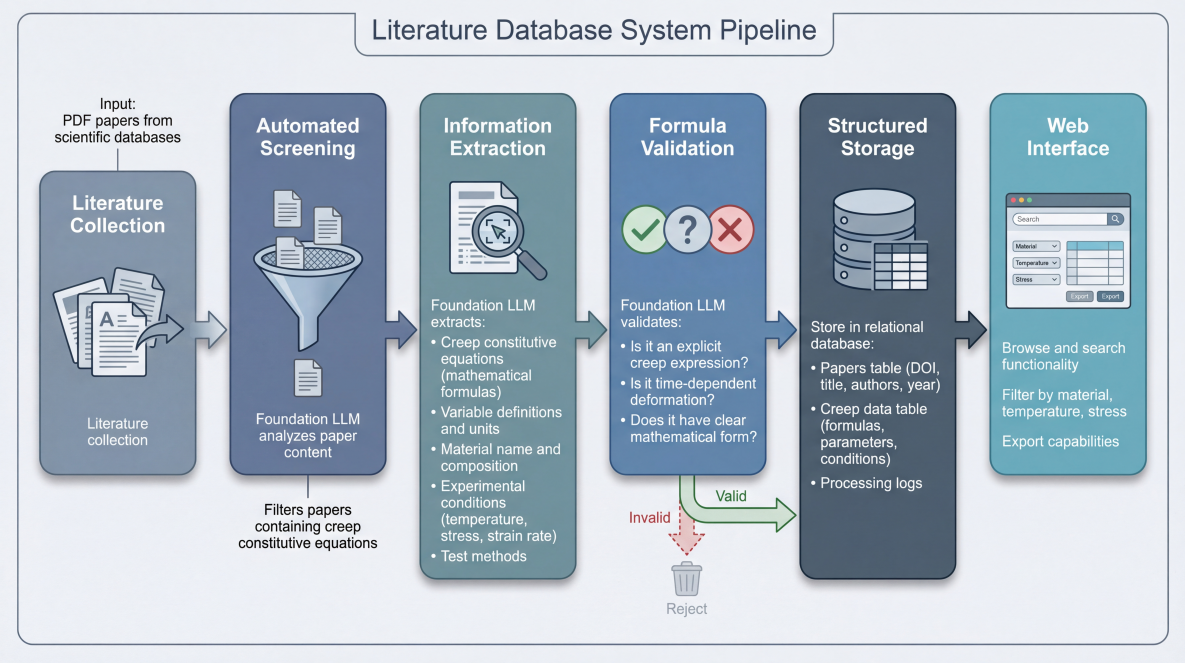

图 1：**自动化文献挖掘流程的架构**。该工作流分为五个不同阶段：(1) **文献收集**：从科学文献库汇集源 PDF；(2) **自动筛选**：基于大语言模型的筛选器分析文档内容，识别包含蠕变本构模型的相关出版物；(3)**信息抽取**：多模态提取引擎解析非结构化文本（提取材料元数据和公式）及图形曲线（提取实验蠕变曲线）；(4)**公式验证**：基于逻辑的验证模块检查所提取表达式的数学完整性和物理合理性；(5)**结构化存储与界面**：验证后的数据被序列化并存入关系数据库，用户可通过易用的 Web 界面进行搜索、筛选和导出。

2.1 文献收集和自动筛选

该流程首先汇集大量科学文献。通过一组与蠕变变形和本构建模相关的全面关键词，从主要科学文献库获取潜在候选文档。

鉴于基于关键词的检索不可避免地会引入噪声，系统采用严格的**自动筛选**模块对语料进行筛选。与传统的基于规则的筛选器不同，该模块利用基础大语言模型的语义理解能力。智能体会分析每篇文档的全文，并根据特定标准判断其相关性：是否包含实验蠕变数据，以及是否讨论本构方程。纯理论性论文（未经实验验证）或与时变形变无关的论文会被自动剔除，从而确保后续处理只针对高价值内容。

2.2 多模态信息提取

系统的核心是提取引擎，它通过处理以下两种不同模态来应对科学文档的异构特性：

- **语义提取（文本与公式）：** 智能体扫描文档，以识别并提取关键元数据，包括材料标识（例如化学成分、标准名称）、实验条件（温度 T、应力 σ）以及蠕变本构定律的显式数学表达式。系统特别关注变量定义和单位标准化，以确保整个数据库的一致性。
- **视觉提取（图形数据）：** 与不加区分的数字化不同，该智能体被设计为优先处理具有代表性的实验轨迹，这些轨迹能够严格映射到文中明确讨论的本构模型。作为文本分析的补充，智能体识别并处理实验图（例如应变—时间曲线），同时过滤掉补充性或含糊的数据序列。它检测坐标系、识别坐标轴刻度，并对相关轨迹中的数据点进行数字化。这一能力使系统能够获取文本中往往缺失、但对模型验证至关重要的原始实验数据。

2.3 物理信息验证协议

为降低生成式人工智能模型常见的“幻觉”风险，工作流程中集成了一个公式验证模块，作为物理信息逻辑门。该模块依据三项标准，对每条提取记录进行严格评估，以确保科学严谨性。具体而言，智能体会核查提取内容是否构成完整、明确的数学方程，而非单纯的描述性片段；检查公式是否具有物理相关性，确保其描述的时变形变符合既有蠕变力学；并通过确认所有变量均有明确定义且单位一致，验证其数学完整性。只有成功通过这项全面验证的记录才会被标记为有效并进入数据库，不符合要求的记录则会被自动标记为拒绝或转交人工审核。

2.4 数据库构建与Web界面

验证后的数据将序列化并存入关系数据库，该数据库包含两个主要实体：用于存储书目元数据（DOI、作者、年份）的论文表，以及用于存储提取出的物理参数和实验条件的蠕变数据表。尤为重要的是，系统保留数字对象唯一标识符（DOI），以确保数据溯源，使每个提取的数据点都能严格追溯至其原始出版物。为方便用户访问，我们在数据库之上开发了易于使用的Web界面。如图2所示，该界面提供浏览和搜索功能。用户可以按材料类型、温度范围或应力水平筛选数据库，并导出结构化数据集以供进一步分析。这一端到端实现不仅能存储提取出的知识，还能使其随时可用于数据驱动的AI4S发现。

2.5 基于技能的认知架构与约束机制

为以高度自主性和可靠性执行上述工作流，我们设计了一种模块化的“基于技能”的认知架构。该框架将复杂的推理任务分解为称为技能的专业化、可执行单元，从而摆脱了单体式智能体设计。形式上，我们将智能体技能（S）建模为认知元组$S = \{I, T, C\}$，其中：

- I（指令注入）：表示特定领域的系统提示，用专家知识“预置”大语言模型，同时明确屏蔽无关的通用知识，以减少噪声。
- T（作用域限定工具）：仅供特定技能使用的一组受限可执行函数。对动作空间进行剪枝，可降低产生幻觉或错误工具调用的概率。
- C（硬约束）：一种强制性的输出模式，用于实施受约束生成，确保输出可供机器执行。

如表1所示，该架构允许智能体在流水线的不同阶段动态实例化专业化角色——从“书目学家”到“物理学家”——并通过实施严格的逻辑门来提高数据保真度。

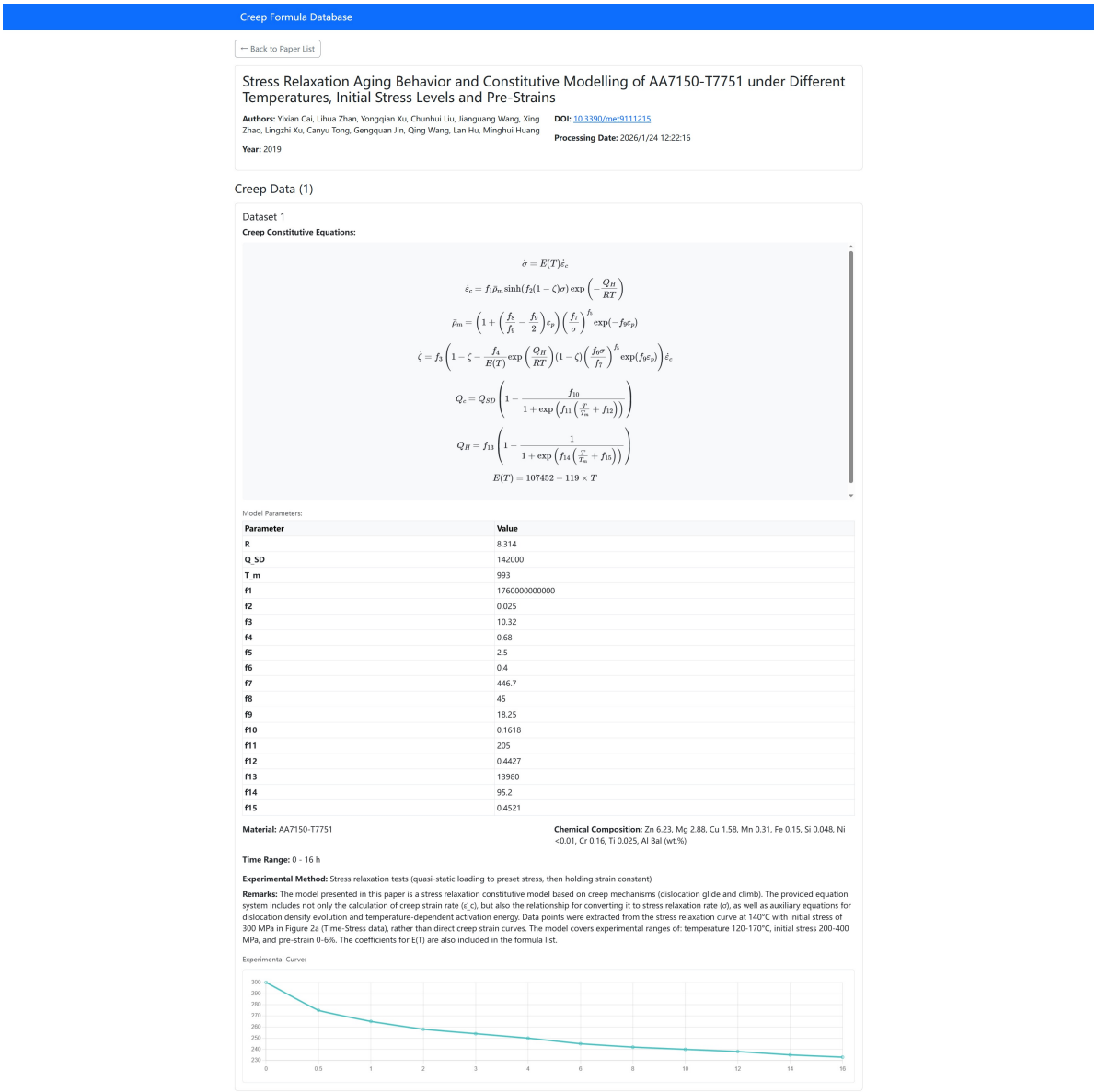

图 2：**AI构建的蠕变数据库的交互式前端界面**。基于网络的平台充当结构化 SQL后端与终端用户之间的桥梁，促进高效的数据检索。主要功能包括：**(1) 精细筛选**（左侧面板）：用户可以通过多维约束查询数据库，例如材料成分、温度范围和应力水平；**(2) 动态可视化**（右侧面板）：所选数据集会立即呈现为应变-时间图，便于快速直观地比较不同实验条件下的蠕变行为；**(3) 数据导出**：经验证的数据点和本构参数可导出为标准化格式，供后续机器学习应用使用。

表1：管道阶段与智能体技能及约束逻辑的映射。

阶段	智能体技能	作用域限定工具(T)	约束机制（C）
收集	文献导航器	基于检索增强生成的检索	**语义消歧**：将有歧义的自然语言描述符转换为精确且符合语法的布尔逻辑；扩展隐含的关键词（例如，“superalloy” $\rightarrow$ “镍基 OR 钴基”）以最大化召回率。
筛选	领域过滤器	零样本分类	**逻辑门控**：实施严格的合取式筛选器（$Pass \iff Has_Data \wedge Has_Equation$）；无论关键词相关性如何，都会自动排除部分匹配（例如，纯模拟ulation studies)（无论关键词相关性如何）。
提取	多模态解析器	视觉语言对齐	**本体与几何对齐**：强制统一的变量模式，以解决命名异质性，同时施加严格的拓扑约束约束（例如单调性），以确保结构完整的数字化轨迹。
验证	物理学护栏	符号推理	**物理信息完整性**：强制遵循量纲齐次定律，并实施严格的统计阈值（$R^2 > 0.9$），用于跨文本参数与视觉数据之间的模态对齐。
存储	数据序列化器	结构化序列化	**模式合规性与溯源**：强制执行严格的数据类型标注，以确保机器可读性，同时嵌入不可变元数据（例如，来源DOI、视觉证据链接）以建立可验证的每条记录的审计追踪。

2.6 自动化工作流程的质量评估

为评估自动筛选模块的效能，我们以经人工验证的真实标准为基准进行了定量评估。从初始语料库中随机选取 20 份文档，组成验证子集。我们对这些文档进行人工审核，并将其归类为“相关”（包含有效的蠕变实验数据）或“不相关”。鉴于此流程属于信息检索任务，我们采用检索增强生成（RAG）系统中使用的标准评估指标来量化性能：精确率、召回率、F1 分数和准确率。为便于不同学科背景的读者理解，我们根据以下分类结果定义这些指标：

- **真正例**（TP）：智能体正确识别出一篇包含蠕变数据的相关论文。
- **假正例**（FP）：智能体错误地将一篇不相关的论文判定为相关（噪声）。
- **真负例**（TN）：智能体正确地拒绝了一篇不相关的论文。
- **假负例**（FN）：智能体未能识别出一篇相关论文（遗漏数据）。

基于这些计数，指标计算如下：**精确率**衡量智能体选择的可靠性，表示检索到的文档中实际相关文档所占的百分比：

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

召回率衡量搜索的完整性，表示智能体成功找到的所有现有相关文档所占的百分比：

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

F1 分数提供是精确率和召回率的调和平均值，能够综合反映智能体的稳健性：

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Fin总体而言，**准确率** 表示分类器在整个验证集上的整体正确性 t:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

这些指标提供了一个全面的框架，用以对智能体在最大化数据获取的同时尽量减少纳入无关文献的能力进行基准测试。

3 结果

为确保所构建数据库的可靠性，我们实施了严格的“人在回路”验证协议。在自动化提取阶段之后，对已处理文献的一个随机子集进行了人工检查。在243份已处理文档中，智能体在正确识别并数字化目标蠕变曲线方面取得了超过90%的经验证成功率。鉴于源文档性质异质，这一高指标尤为重要：这些文档既包括高分辨率的现代数字PDF，也包括跨越数十年的扫描早期出版物，后者往往清晰度不一，并带有栅格化伪影。图3给出了定性比较，凸显了智能体在处理复杂提取场景时的稳健性。自动化文献挖掘的一大挑战是视觉信息密度高；单篇出版物通常包含多种实验可视化图——从拉伸应力-应变曲线到显微组织显微图——必须将其与目标蠕变数据区分开来。此外，即便在一幅已识别的蠕变图中，也常常并存多条轨迹，分别代表不同的应力或温度条件（如图3a所示，52.7 MPa与31.6 MPa的曲线相互重叠）。如重建结果（图3b）所示，智能体成功执行了语义解缠：它不仅滤除了网格线和背景噪声等图形伪影，还准确分离出与目标实验条件（$\sigma = 31.6$ MPa）相对应的特定轨迹，从而有效忽略了无关数据序列。这表明一种超越简单视觉分割的情境感知提取能力。除图形数据之外，所提取元数据的语义准确率——具体包括实验条件（温度、应力）和材料标识符——也得到了验证。这些文本提取的一致性可由先前讨论的参数空间中观察到的逻辑聚类加以证明，其中未检测到明显的物理异常值（例如冰出现在高温聚类中）。对视觉（曲线）与语义（文本）信息的双重验证证实，该自动化工作流已足够精确，可作为数据驱动的科学探究的基础。

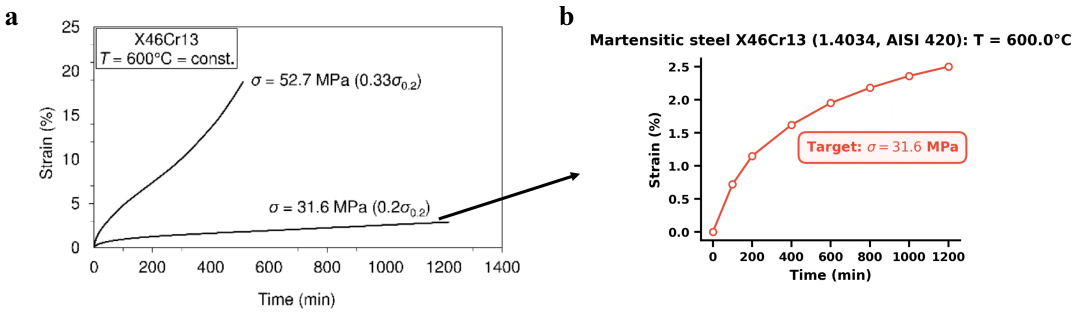

图3：**复杂多变量图中选择性曲线提取示例**。(a)原始图像（材料：X46Cr13，T=600°C），其中包含不同应力水平[32]下的多条蠕变轨迹。(b)根据智能体输出重建的数字化结果。智能体成功分离了视觉信息，准确提取出与目标条件（$\sigma = 31.6$ MPa）对应的特定曲线，同时忽略了无关的数据序列（52.7 MPa）。值得注意的是，提取的数据点具有很高的保真度，保留了潜在的蠕变趋势。

3.1 构建的数据库概述与组成

为评估所提出的Agent-Skills框架的有效性，我们部署该系统，对非结构化学术文献的收集执行自动化提取任务。该智能体处理了243篇出版物，并成功构建了353个高保真的“方程-数据”对。虽然单篇出版物可能包含多个图形数据集，

我们的框架侧重于分离原作者用于推导或验证其本构模型的主要轨迹。这种严格筛选产生了一个聚焦的数据集，其中每条曲线都附带其显式数学参数，优先考虑物理可解释性而非原始数据量。随后对所提取数据的人工验证表明，自动化工作流具有高保真度，经验证的成功率超过90%。这一高精确率凸显了所定义技能在处理不同期刊及数十年研究中异构格式与命名方面的稳健性。

所得数据库在材料成分上呈现出显著的多样性，表明该智能体能够在不同科学子领域之间进行泛化。如图4（材料分布）所示，该数据集主要由高性能工程材料构成，其中镍基合金（19.0%）和钢/铁（12.5%）占比最大。这些是高温应用的关键材料，反映了文献对工业相关性的关注。值得注意的是，该智能体还成功检索到了较为非常规类别的数据，包括聚合物/塑料（8.5%）、岩石/地质材料（5.9%），甚至冰/冰川（4.0%）。这种对多样材料类别的纳入——从软物质到地质复合材料——验证了该智能体并未过度拟合于特定金属关键词，而是对广义定义的“蠕变”现象具备语义理解。

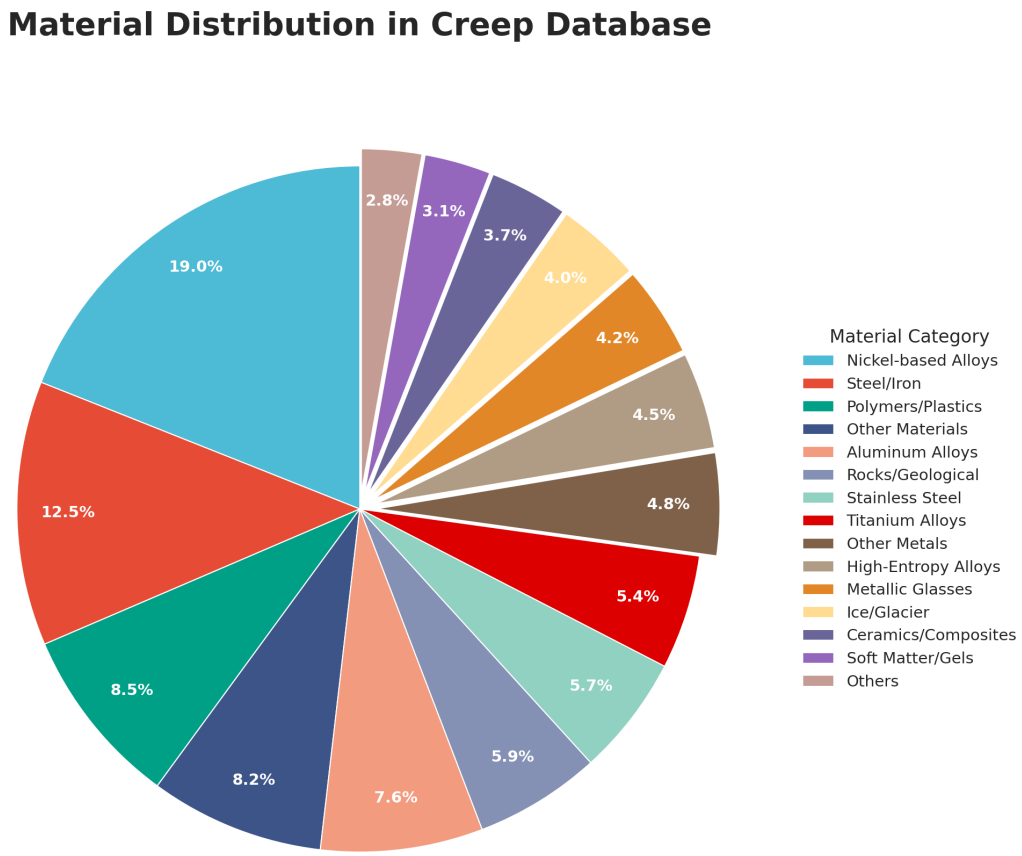

图4：**所构建蠕变数据库中材料成分的多样性**。 该扇形图展示了353条已提取蠕变曲线中材料类别的分布。该数据集展现出广泛的领域覆盖范围，既涵盖关键工业材料（例如镍基合金，19.0%；钢/铁，12.5%），也包括非金属体系（例如聚合物，8.5%；冰/冰川，4.0%）。这种多样性证实了该智能体具备语义能力，能够在异质的材料科学子领域中识别相关的“蠕变”行为。

此外，该数据库涵盖了全面的实验条件范围，为未来建模提供了丰富的参数空间。图5可视化了关键物理参数的分布：温度和应力。温度分布（图5a）范围从低温和室温环境（典型见于聚合物

和冰）到超过1000°C的极端条件（超合金和陶瓷的典型特征）。同样，应力水平（图5b）范围从低量级地质压力到高应力工程载荷（>1000 MPa）。这些变量之间的相互作用被映射在图6所示的温度-应力参数空间中。该散点图揭示了对应于不同材料性能的明显聚类——例如，超合金所占据的高温/高应力区域，与软物质的低温/低应力区域。这一广阔物理空间的成功填充证实，所构建的数据库不仅规模可观，而且具有物理代表性，有效捕捉了材料在多个尺度上的本构行为。

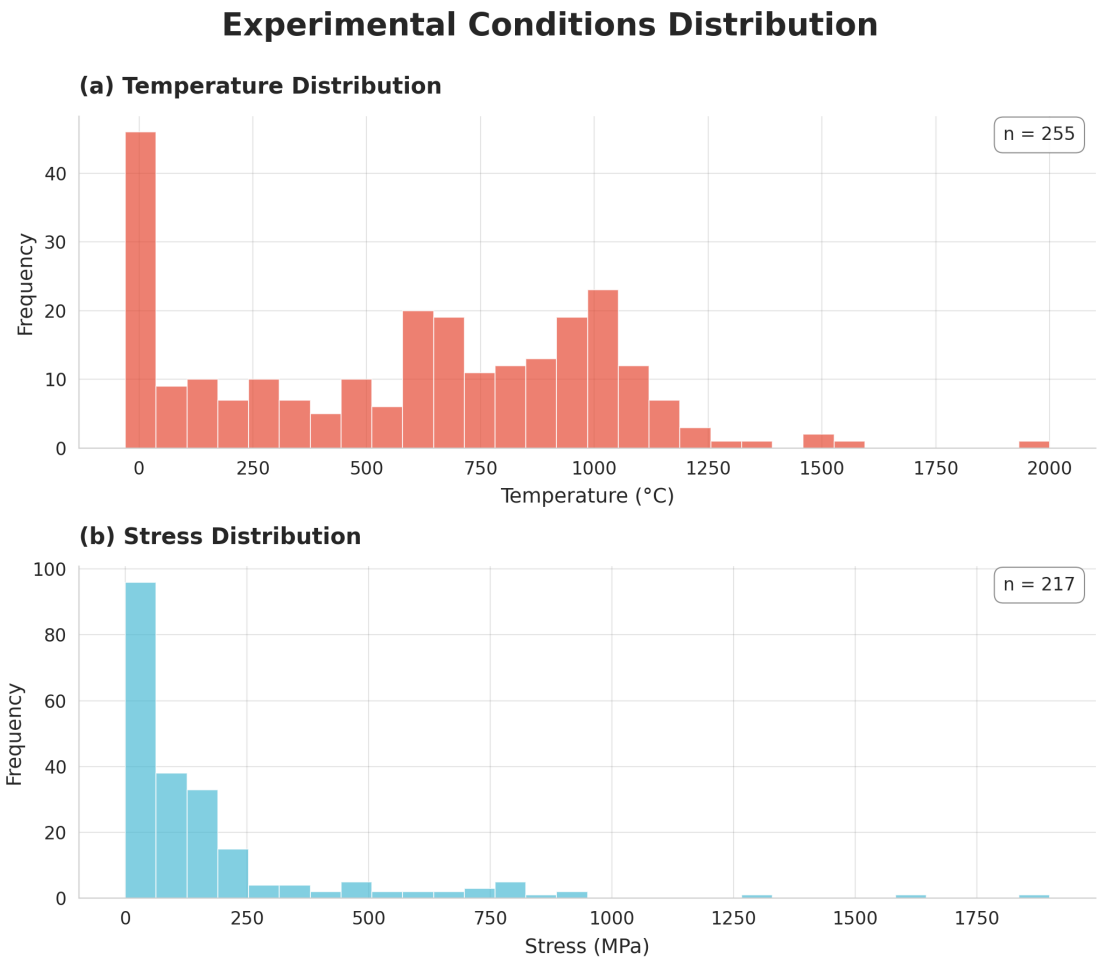

图5：**从文献中提取的实验条件统计分布**。(a) 实验温度（°C）的频率分布，范围涵盖低温环境至超高温区域（>1000°C）。(b) 施加应力水平（MPa）的频率分布。运行条件范围广泛，表明该数据库全面呈现了历史文献中记录的材料性能边界。

3.2 通过跨模态对齐验证物理一致性

除了图形准确率之外，科学数据库的一项关键要求是保持所提取的原始数据与源文本中描述的理论模型之间的物理一致性。科学出版物中经常存在信息冗余：既以图形形式呈现实验结果，同时又在随附文本中报告拟合本构参数。这为跨模态验证提供了独特的机会。为此，我们分析了一个关于 AMAG-183 金属玻璃的代表性案例研究（图7）。智能体独立执行了两项任务：（1）从背景带有颜色且包含多个数据序列的复杂图表中数字化提取原始变形-时间轨迹（图7a）；（2）从文档的讨论部分提取数学建模参数。在这个具体案例中，源文献将变形行为描述为

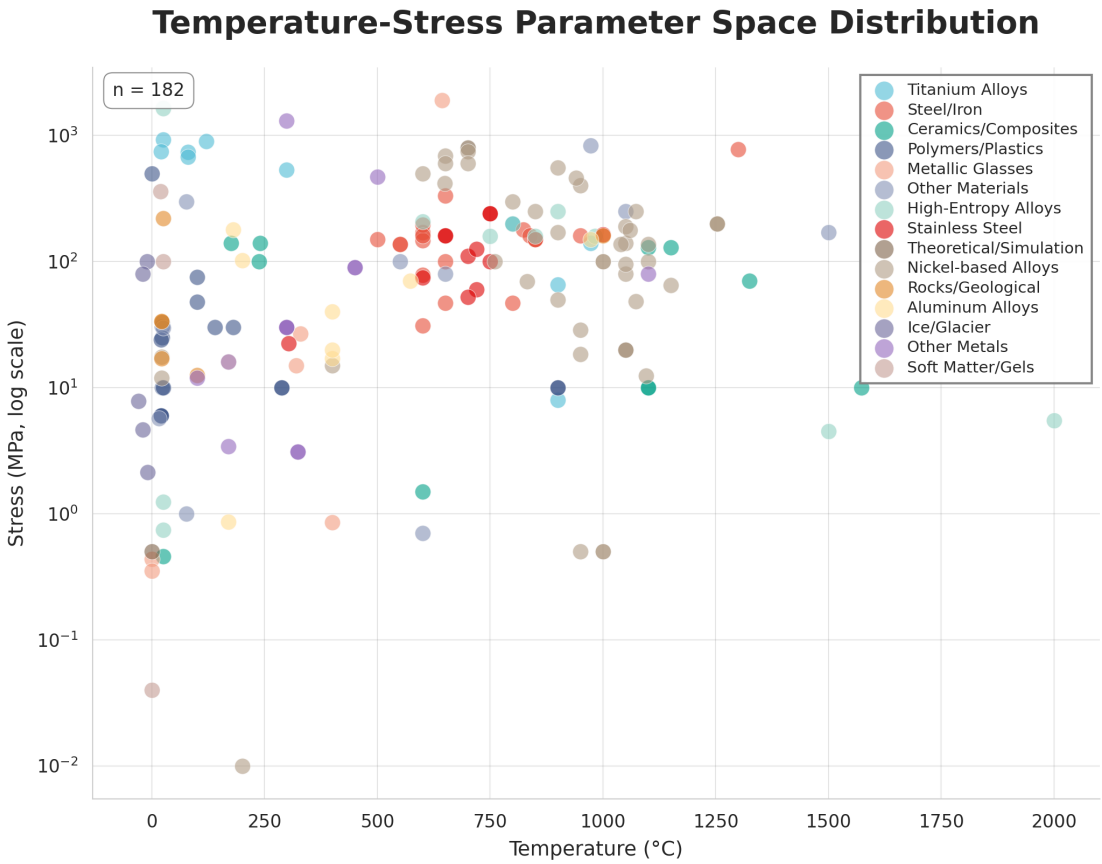

图 6：**温度-应力参数空间的可视化**。 按材料类别着色的散点图，展示了所有提取曲线（$n = 353$）对应的实验条件。该可视化呈现出与材料性能相对应的明显物理簇，例如超合金主要分布在高温、高应力区域，而软物质则处于低能量范围，这表明自动构建的数据集具有物理一致性和合理的分布。

采用非线性 Duffing振子模型来捕捉剪切带演化的混沌动力学。图7b展示了这两种模态之间的对齐情况。空心圆表示通过视觉方式数字化提取的实验数据，红色实线则表示将文本中提取的参数代入控制微分方程后重建的理论轨迹。根据插图中的量化结果，理论重建模型与数字化提取的数据点高度吻合，决定系数（R^2）达到 0.9999。这种近乎完美的同步表明，智能体有效弥合了非结构化文本参数与栅格化视觉数据之间的语义鸿沟，确保数据库即使针对复杂的非线性材料行为，也能保持物理自洽。这一严格的跨模态验证协议还起到了隐式质量筛选的作用；与不加甄别地提取相比，它可能会限制每篇出版物中提取曲线的总数，但能确保数据库中的每条记录都具备可验证的高保真联系，将实验数据与物理理论关联起来。

4 结论

本研究提出了一种新颖的自主框架，能够将零散的科学文献转化为结构化、可由机器执行的知识。通过将基础大语言模型的推理能力与精准的视觉提取技能相结合，我们成功突破了长期阻碍力学和材料科学数据驱动研究的“数据瓶颈”。蠕变力学数据库的构建为这一方法提供了严格的试验平台。结果表明，该智能体超越了传统基于光学字符识别的方法所面临的局限，展现出必要的语义智能，能够解析复杂的实验图，并依据物理定律验证数据。在我们的案例研究中，数字化曲线与

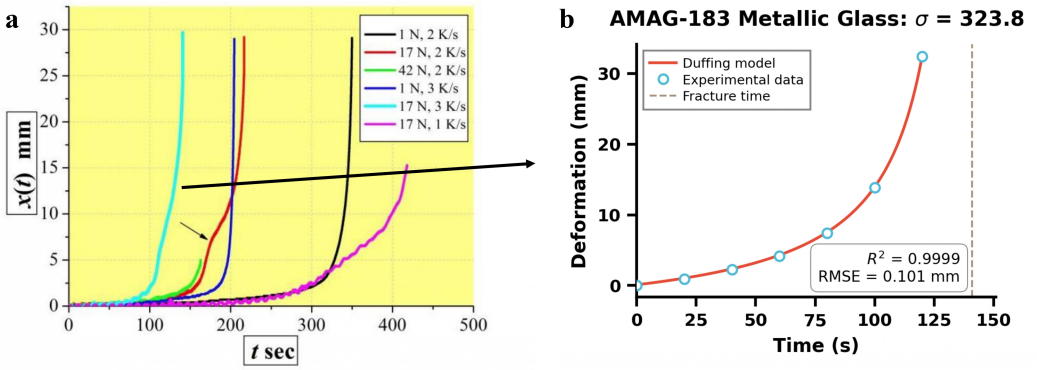

图7：**复杂非线性行为物理一致性的跨模态验证**。(a) AMAG-183 金属玻璃的原始实验图，其背景非标准且包含多组重叠数据序列[33]。(b) 将提取的数据与理论物理模型进行验证。符号（蓝色圆圈）表示从图像中数字化提取的数据，实线（红色）表示使用从文本中提取的*Duffing*模型参数计算出的轨迹。极高的相关性（$R^2 = 0.9999$）展现了智能体将文本数学描述正确映射到相应视觉表示的能力，从而验证了所构建数据库的物理完整性。

理论模型之间近乎完美的相关性（{v1}），凸显了该系统进行跨模态推理的能力——有效弥合视觉观察与数学描述之间的差距。尽管本研究聚焦于蠕变变形，但其底层架构不受特定领域限制。“智能体-技能”框架的模块化特性使其能够快速适配其他专业领域，从电池循环性能到催化反应动力学。随着大语言模型不断发展，我们设想此类自主数据智能体将成为科学基础设施的标准组成部分，使研究人员摆脱繁琐的数据整理工作，并在人工智能驱动的科学时代加速假设生成与发现的循环。

参考文献

[1] Hanchen Wang、Tianfan Fu、Yuanqi Du、Wenhao Gao、Kexin Huang、Ziming Liu、Payal Chandak、Shengchao Liu、Peter Van Katwyk、Andreea Deac 等. 人工智能时代的科学发现. *Nature*, 620(7972):47–60, 2023.[2] Ankit Agrawal 和 Alok Choudhary. 观点：材料信息学与大数据：在材料科学中实现科学的“第四范式”. *Apl Materials*, 4(5), 2016.[3] Mark Chen. 对代码训练的大语言模型进行评估. *arXiv preprint arXiv:2107.03374*, 2021.[4] Sébastien Bubeck、Varun Chandrasekaran、Ronen Eldan、Johannes Gehrke、Eric Horvitz、Ece Kamar、Peter Lee、Yin Tat Lee、Yuanzhi Li、Scott Lundberg 等. 通用人工智能的火花：对 GPT-4 的早期实验. *arXivpreprint arXiv:2303.12712*, 2023.[5] Jason Wei、Xuezhi Wang、Dale Schuurmans、Maarten Bosma、Fei Xia、Ed Chi、Quoc V Le、Denny Zhou 等. 思维链提示能够激发大语言模型的推理能力. *Advances in neural informationprocessingsystems*, 35:24824–24837, 2022.[6] Shunyu Yao、Jeffrey Zhao、Dian Yu、Nan Du、Izhak Shafran、Karthik R Narasimhan 和 Yuan Cao. ReAct：协同语言模型中的推理与行动. 载于*The eleventhinternational conference on learningrepresentations*, 2022.[7] Timo Schick、Jane Dwivedi-Yu、Roberto Dessì、Roberta Raileanu、Maria Lomeli、Eric Hambro、Luke Zettlemoyer、Nicola Cancedda 和 Thomas Scialom. Toolformer：语言模型可以自学使用工具。*Advances in Neural Information Processing Systems*, 36:68539–68551, 2023.[8] Zhiheng Xi、Wenxiang Chen、Xin Guo、Wei He、Yiwen Ding、Boyang Hong、Ming Zhang、Junzhe Wang、Senjie Jin、Enyu Zhou 等. 基于大语言模型的智能体的兴起与潜力：综述. *Science ChinaInformationSciences*, 68(2):121101, 2025.

[9] Daniil A Boiko、Robert MacKnight、Ben Kline 和 Gabe Gomes. 利用大语言模型开展自主化学研究. *Nature*, 624(7992):570–578, 2023.[10] Nathan J Szymanski、Bernardus Rendy、Yuxing Fei、Rishi E Kumar、Tanjin He、David Milsted、Matthew J McDermott、Max Gallant、Ekin Dogus Cubuk、Amil Merchant 等. 用于加速合成新型材料的自主实验室. *Nature*, 624(7990):86–91, 2023.[11] Amil Merchant、Simon Batzner、Samuel S Schoenholz、Muratahan Aykol、Gowoon Cheon 和 Ekin Dogus Cubuk. 扩展深度学习以推动材料发现. *Nature*, 624(7990):80–85, 2023.[12] Zhuo Wang、Zhehao Sun、Hang Yin、Xinghui Liu、Jinlan Wang、Haitao Zhao、Cheng Heng Pang、Tao Wu、Shuzhou Li、Zongyou Yin 等. 数据驱动的材料创新与应用. 《先进材料》, 34(36):2104113, 2022.[13] Vincent Fan、Yujie Qian、Alex Wang、Amber Wang、Connor W Coley 和 Regina Barzilay. Openchemie：用于化学文献信息抽取的工具包. 《化学信息与建模杂志》, 64(14):5521–5534, 2024.[14] Matthew C Swain 和 Jacqueline M Cole. Chemdataextractor：用于从科学文献中自动抽取化学信息的工具包. 《化学信息与建模杂志》, 56(10):1894–1904, 2016.[15] Leigh Weston、Vahe Tshitoyan、John Dagdelen、Olga Kononova、Amalie Trewartha、Kristin A Persson、Gerbrand Ceder 和 Anubhav Jain. 将命名实体识别与规范化应用于从材料科学文献中大规模提取信息. 《化学信息与建模杂志》, 59(9):3692–3702, 2019.[16] Zheren Wang、Olga Kononova、Kevin Cruse、Tanjin He、Haoyan Huo、Yuxing Fei、Yan Zeng、Yingzhi Sun、Zijian Cai、Wenhao Sun 等. 从科学文献中提取的基于溶液的无机材料合成流程数据集. 科学数据, 9(1):231, 2022.[17] Ying Zhang 和 Chen Ling. 将机器学习应用于材料科学小数据集的策略. 《*npj*计算材料学》, 4(1):25, 2018.[18] Andres M Bran、Sam Cox、Oliver Schilter、Carlo Baldassari、Andrew D White 和 Philippe Schwaller. ChemCrow：利用化学工具增强大语言模型. *arXivpreprint arXiv:2304.05376*, 2023.[19] Vahe Tshitoyan、John Dagdelen、Leigh Weston、Alexander Dunn、Ziqin Rong、Olga Kononova、Kristin A Persson、Gerbrand Ceder 和 Anubhav Jain. 无监督词嵌入从材料科学文献中捕获潜在知识. *Nature*, 571(7763):95–98, 2019.[20] Junyu Luo、Zekun Li、Jinpeng Wang 和 Chin-Yew Lin. ChartOCR：通过深度混合框架从图表图像中提取数据. 载于 *IEEE/CVF* 冬季计算机视觉应用会议论文集 , 第 1917–1925 页 , 2021.[21] Ahmed Masry 、 Xuan Long Do、Jia Qing Tan、Shafiq Joty 和 Enamul Hoque. ChartQA：用于图表问答的基准测试，结合视觉与逻辑推理. 载于计算语言学协会：*ACL 2022*论文集, 第2263–2279页, 2022.[22] Ziwei Ji、Nayeon Lee、Rita Frieske、Tiezheng Yu、Dan Su、Yan Xu、Etsuko Ishii、Ye Jin Bang、Andrea Madotto 和 Pascale Fung. 自然语言生成中的幻觉综述 . 《 *ACM* 计算评论》 , 55(12):1–38, 2023.[23] Patrick Lewis 、Ethan Perez、Aleksandra Piktus、Fabio Petroni、Vladimir Karpukhin、Naman Goyal、Heinrich Küttler、Mike Lewis、Wen-tau Yih、Tim Rocktäschel 等. 面向知识密集型自然语言处理任务的检索增强生成. *Advances inneuralinformation processing systems*, 33:9459–9474, 2020.[24] Haotian Liu、Chunyuan Li、Qingyang Wu 和 Yong Jae Lee. 视觉指令微调. *Advancesinneuralinformation processing systems*, 36:34892–34916, 2023.[25] Michael E Kassner. 金属和合金蠕变基础. Butterworth-Heinemann, 2015.[26] Kota Sawada、Kazuhiro Kimura、Fujio Abe、Yasushi Taniuchi、Kaoru Sekido、Takehiro Nojima、Toshio Ohba、Hideaki Kushima、Hideko Miyazaki、Hiromichi Hongo 和 Takashi Watanabe. NIMS 蠕变数据表目录.《先进材料科学与技术》, 20(1):1131–1149, 2019.[27] Silviu-Marian Udrescu 和 Max Tegmark. AI Feynman：一种受物理学启发的符号回归方法.《科学进展》, 6(16):eaay2631, 2020.[28] George Em Karniadakis、Ioannis G Kevrekidis、Lu Lu、Paris Perdikaris、Sifan Wang 和 Liu Yang. 物理学启发的机器学习. 《自然评论：物理学》, 3(6):422–440, 2021.

[29] Maziar Raissi、Paris Perdikaris 和 George E Karniadakis。《物理信息神经网络：一种用于求解涉及非线性偏微分方程的正向和反向问题的深度学习框架》。*Journal ofComputational physics*，378：686–707，2019。[30] Xiao Zhang 和 Shunbo Hu。《一种用于物理信息神经网络的多任务动态互联学习方法》。*Materials TodayCommunications*，第 113054 页，2025。[31] An Yang、Anfeng Li、Baosong Yang、Beichen Zhang、Binyuan Hui、Bo Zheng、Bowen Yu、Chang Gao、Chengen Huang、Chenxu Lv 等。《Qwen3 技术报告》。*arXiv* 预印本 *arXiv:2505.09388*，2025。[32] Josip Brnic、Sanjin Krscanski、Domagoj Lanc、Marino Brcic、Goran Turkalj、Marko Canadija 和 Jitai Niu。《马氏体钢 X46Cr13 的力学行为、抗蠕变性能和单轴疲劳强度分析》。*Materials*，10(4)，2017。[33] AD Berezner、VA Fedorov、NS Perov 和 GV Grigoriev。《杜芬方程的新型精确解：稳定性分析及其在实际非线性变形试验中的应用》。*arXiv* 预印本 *arXiv:2512.24774*，2025。