arXiv:2601.05051v1 [cs.AI] 8 Jan 2026

在材料科学领域发表 FAIR（可发现、可访问、可互操作、可重用）且机器可操作的综述：神经符号人工智能中符号知识的价值

Jennifer D'SouzaTIB 莱布尼茨科学与技术信息中心，德国汉诺威 jennifer.dsouza@tib.eu

Sören AuerTIB 莱布尼茨科学与技术信息中心，德国汉诺威；L3S 研究中心，汉诺威莱布尼茨大学，德国汉诺威

Eleni Poupaki埃因霍温理工大学（TU/e），荷兰埃因霍温

Alex Watkins华威大学，英国考文垂

Anjana Devi莱布尼茨固态与材料研究所（IFW），德国德累斯顿

Riikka L. Puurunen阿尔托大学化学工程学院，芬兰埃斯波

Bora Karasulu华威大学，英国考文垂

Adrie Mackus埃因霍温理工大学（TU/e），荷兰埃因霍温

Erwin Kessels埃因霍温理工大学（TU/e），荷兰埃因霍温

摘要

科学综述是材料科学知识整合的核心，但其关键洞见仍被锁定在叙述性文本和静态 PDF 表格中，限制了人类和机器对其进行再利用。本文以原子层沉积与刻蚀（ALD/E）为案例，介绍我们如何将综述表格作为 FAIR（可发现、可访问、可互操作、可重用）的机器可操作比较发布到开放研究知识图谱（ORKG）中，将其转化为结构化、可查询的知识。在此基础上，我们比较了在 ORKG 上进行的符号查询与基于大语言模型的查询，并主张精选符号层应继续作为材料科学领域可靠的神经符号人工智能的骨干，而大语言模型则应作为以符号为基础的互补接口，而非独立的真值来源。

预印本。正在审稿。

> “设想一种未来的设备……其中，一个人存储了自己的全部书籍、记录和通信，并且该设备经过机械化处理，可以极快且灵活地供人查阅。它是对其存储器的一种扩展性私密补充。”
>
> — 范内瓦·布什，1945年7月，《诚如所思》 [9]

1 引言

范内瓦·布什于1945年提出的个人化、可搜索的“memex”愿景[9]与当今材料科学领域的情形十分契合。原子层沉积与刻蚀（原子层沉积/原子层刻蚀，ALD/ALE）如今产生了海量关于工艺、机理和性能数据的文献，其中许多内容通过叙述性文本和多页PDF表格传达。综述文章通过汇总这些信息提供了帮助，但要回答一个简单的比较问题（例如，“针对材料X，哪些已报道的原子层沉积工艺满足指定的温度范围和每周期生长量区间？”），通常仍需付出大量人工工作——阅读文献、在电子表格中重建表格并筛选条目。因此，提取出的综合信息往往以临时格式保存（例如个人笔记或本地电子表格），难以分享、查询和系统性地重用。

与此同时，人们越来越关注如何使科学知识具备机器可操作性：按照一定结构组织知识，使其能够被自动搜索、筛选和重新组合，这与科学数据的FAIR指导原则（可发现、可访问、可互操作、可重用）相一致 [62]。开放研究知识图谱（ORKG）[6, 7] 便是这样一种新一代基础设施。作者和策展人不再仅仅发表PDF，而是可以将关键发现表示为结构化的“比较”，其形式类似常见表格——每行对应一项研究，每列对应材料、前驱体、基底、工艺条件和性能指标——但这些内容存储为带有稳定标识符和显式模式的知识图谱。早期研究已展示，此类“SmartReviews”能够使调查类文章中的表格同时具备人类可读性和机器可读性[46], ; MatKG[61] 等相关工作则表明，基于文献构建大规模知识图谱在材料科学领域是可行的。

与此同时，大语言模型（LLM）已迅速成为材料研究人员青睐的工具。例如，近期有关原子层沉积基准测试的研究评估了 GPT-4级模型在原子层沉积工艺设计和概念问题上的辅助能力[65], ; 而 MOF-ChemUnity 则将文献衍生知识库与基于大语言模型的辅助功能相结合，用于金属有机框架研究[53]。这些系统之所以引人注目，是因为它们能够阅读 PDF 文件、用自然语言回答问题，并帮助解释趋势或提出假设。然而，大语言模型也是随机生成模型：它们从可能的下一个词元的概率分布中抽样，因此同一个问题在不同运行中可能得到略有不同的答案。当精确数值和准确的表格条目至关重要时——例如工艺窗口、每周期生长量或刻蚀选择性——这种缺乏确定性的特点，加上幻觉事实和误读表格等已知问题，限制了它们作为定量决策独立工具的适用性[69, 18]。

在这项工作中，我们针对 ALD/ALE 综述文章这一特定场景，将这两项进展结合起来。我们的出发点是构建符号层：选取九篇近期综述论文（其中三篇关于 ALD，六篇关于 ALE），并将其中 18 个符合条件的综述表格——即总结多项原始研究中的工艺及条件的表格——转换为机器可操作的 ORKG 比较。每个比较都是经过人工整理的结构化表格，包含材料、前驱体化学、基底、工艺条件和性能描述符等类型化列，且每一行都链接回所引用的文献。在此符号表示的基础上，我们进一步研究当前大语言模型在回答有关这些表格的精确科学问题时表现如何，既考察它们只能查看原始 PDF 文件的情况，也考察它们明确以 ORKG 比较为依据的情况。在这些设置下，我们发现，对于综述类问题，基于 ORKG 比较进行符号查询能够提供最精确、可复现的答案，而仅基于 PDF 的大语言模型查询则常常难以处理表格结构和数值准确性。专家反馈证实，这些机器可操作的比较对于 ALD/ALE 研究中的比较推理具有直接实用价值。以这些结构化表格为大语言模型提供依据，能显著提升答案可靠性，但仍无法完全取代作为金标准的符号查询。

具体而言，我们探讨三个研究问题 ns:

RQ1 *ALD/ALE* 专家认为，机器可操作的 *ORKG* 比较表在回答复杂综述问题时有多大用处？我们将 ORKG 表格视为一种新型综述成果，并请在职研究人员评价其有意义性和实用性，同时请他们评论其局限性以及希望支持的查询。

研究问题2 仅基于综述文章 *PDF* 运行的大语言模型，在多大程度上能够重现通过符号查询获得的精确表格形式答案？在此，我们将 PDF 文件上传至最先进的大语言模型，提出与 SPARQL 查询对应的详细问题，并衡量模型生成的表格与基于 ORKG 的金标准之间的匹配程度。

RQ3 当大语言模型明确以机器可操作的 *ORKG* 比较表为依据时，回答相同问题的效果如何？与符号化 *SPARQL* 基线以及纯粹基于 *PDF* 的查询相比，又表现如何？在这种设置下，模型会接收准确的 ORKG 比较（导出为 CSV 表格）以及自然语言问题，使我们能够检验结构化符号上下文是否能缩小与符号查询之间的部分差距。

我们的贡献有三点。第一，我们发布了来自九篇原子层沉积/原子层刻蚀综述文章的18个精选 ORKG 比较表，以及33个自然语言问题及其 SPARQL 公式，并将其作为*ALD/E* 查询数据集[15]发布。第二，我们报告了一项领域专家评估：三位原子层沉积/原子层刻蚀研究人员认为，这些问题和由 SPARQL 得出的表格具有科学意义，并且对比较推理具有实际用途。第三，我们针对相同内容，对三种问答模式进行了基准测试——(i) 在 ORKG 上使用符号 SPARQL，(ii) 仅基于 PDF 的大语言模型查询，以及 (iii) 基于 ORKG 表格的大语言模型查询——并指出了各自的优势与局限。我们希望向 JVSTA 读者传达的核心信息是：除 PDF 外，以符号化、机器可操作的形式发布综述知识，既能支持当下可靠的专家查询，也能为未来更可信的神经符号助手奠定基础，助力原子层沉积/原子层刻蚀工作流程。

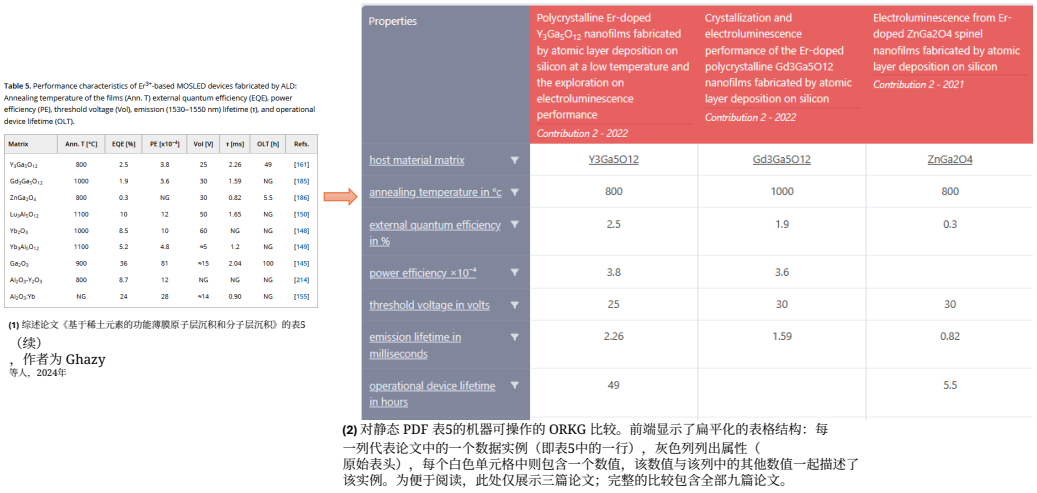


Table 5. Performance characteristics of Er^{3+}-based MOSLED devices fabricated by ALD: Annealing temperature of the films (Ann. T) external quantum efficiency (EQE), power efficiency (PE), threshold voltage (Vol), emission (1530–1550 nm) lifetime (τ), and operational device lifetime (OLT).

Matrix	Ann. T [°C]	EQE [%]	PE [x10^{-4}]	Vol [V]	τ [ms]	OLT [h]	Refs.
$Y_3Ga_5O_{12}$	800	2.5	3.8	25	2.26	49	[161]
$Gd_3Ga_5O_{12}$	1000	1.9	3.6	30	1.59	NG	[185]
$ZnGa_2O_4$	800	0.3	NG	30	0.82	5.5	[186]
$Lu_3Al_5O_{12}$	1100	10	12	50	1.65	NG	[150]
Yb_2O_3	1000	8.5	10	60	NG	NG	[148]
$Yb_3Al_5O_{12}$	1100	5.2	4.8	≈5	1.2	NG	[149]
Ga_2O_3	900	36	81	≈15	2.04	100	[145]
Al_2O_3-Y_2O_3	800	8.7	12	NG	NG	NG	[214]
Al_2O_3:Yb	NG	24	28	≈14	0.90	NG	[155]

(1) 综述论文《基于稀土元素的功能薄膜原子层沉积和分子层沉积》的表5（续），作者为 Ghazy 等人，2024年

Properties	Polycrystalline Er-doped $Y_3Ga_5O_{12}$ nanofilms fabricated by atomic layer deposition on silicon at a low temperature and the exploration on electroluminescence performance *Contribution 2 - 2022*	Crystallization and electroluminescence performance of the Er-doped polycrystalline Gd3Ga5O12 nanofilms fabricated by atomic layer deposition on silicon *Contribution 2 - 2022*	Electroluminescence from Er-doped ZnGa2O4 spinel nanofilms fabricated by atomic layer deposition on silicon *Contribution 2 - 2021*
host material matrix	Y3Ga5O12	Gd3Ga5O12	ZnGa2O4
annealing temperature in °c	800	1000	800
external quantum efficiency in %	2.5	1.9	0.3
power efficiency ×10^{-4}	3.8	3.6	
threshold voltage in volts	25	30	30
emission lifetime in milliseconds	2.26	1.59	0.82
operational device lifetime in hours	49		5.5

(2) 对静态 PDF 表5的机器可操作的 ORKG 比较。前端显示了扁平化的表格结构：每一列代表论文中的一个数据实例（即表5中的一行），灰色列列出属性（原始表头），每个白色单元格中则包含一个数值，该数值与该列中的其他数值一起描述了该实例。为便于阅读，此处仅展示三篇论文；完整的比较包含全部九篇论文。

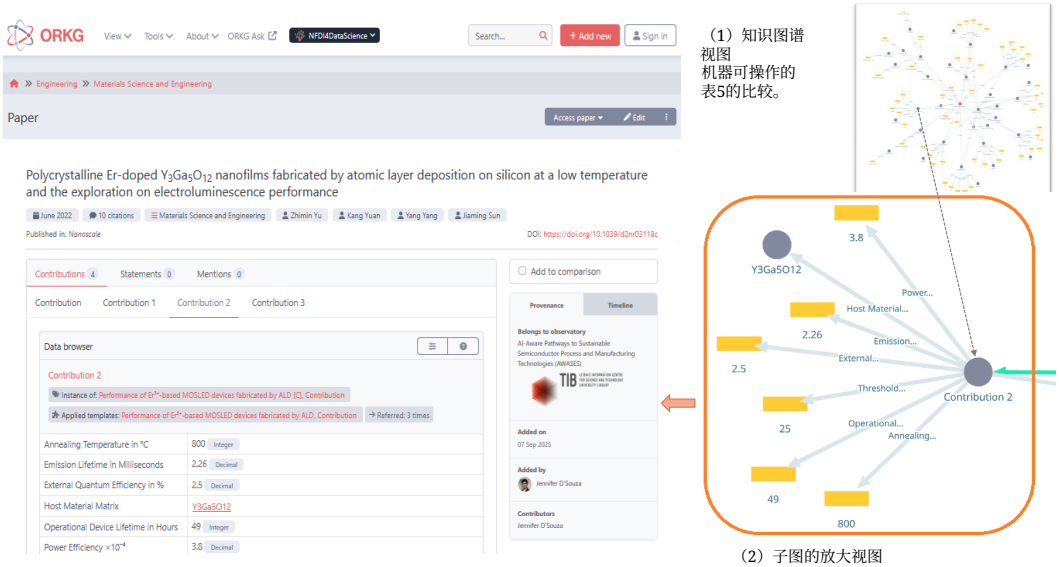
图1：将综述表格建模为开放研究知识图谱（ORKG）中的机器可操作数据。面板（1）展示了 Ghazy等人 [25]的稀土原子层沉积综述中的表5。面板（2）展示了对应的 ORKG 比较：https://orkg.org/comparisons/ R1469991。原表格中的各行被表示为贡献列，原表格的列标题则作为 ORKG 属性。为便于阅读，仅展示了完整比较中九篇论文里的三篇。以下七项属性被建模为网络资源：主体材料基质、以 ℃为单位的退火温度、以%为单位的外量子效率、功率效率 $\times 10^{-4}$、以伏特为单位的阈值电压、以毫秒为单位的发射寿命，以及以小时为单位的器件工作寿命。

2 背景：开放研究知识图谱

大型工业知识图谱——例如 Google、Microsoft、Facebook、eBay 和 IBM 开发的知识图谱——展示了将信息组织为实体 和关系 如何改变人们获取知识

[57, 42]。在这些场景中，电影、人物或产品等项目被建模为节点（例如某部电影、一位演员或一部智能手机），而它们的属性和连接（如“由……执导”“与……共同出演”“价格为”或“由……制造”）则表示为带标签的边。这种简单模式——一组可比较的项目，每个项目都按照一组共同属性进行描述——是许多面向用户的服务的基础：由于相关事实以一致的机器可读结构存储，而非埋藏在自由文本中，人们可以快速检索“2010年之后由 X 执导的所有电影”或“价格低于 Z 且具备 Y 特性的所有产品”。

开放研究知识图谱（ORKG） [6, 7], 也遵循类似的原则。它是一个在线平台（ https://orkg.org/），以相同方式组织科学文献：将研究论文、材料、工艺和实验条件视为实体，并通过明确定义的属性来记录它们之间的关系。静态 PDF 表格中，诸如“基质”“退火温度”或“阈值电压”等列标题可能在不同论文中反复出现，却彼此孤立、无法互操作；与之不同，ORKG 将这些内容建模为明确且可搜索的属性。“前驱体”“共反应物”或“每周期生长量”等关系则成为可在不同作品间对齐和复用的命名谓词 [16, 26]。



（1）知识图谱视图
机器可操作的表5的比较。

（3）结构化工艺描述的 ORKG 论文前端视图。

（2）子图的放大视图
对应于一项工艺描述。

图2：图1的延续。从面板（1）开始，即知识图谱视图（机器可操作表格 5 [25] 的实际图表示）。面板2放大了第一项贡献，将比较条目与 ORKG 中对应的论文记录关联起来：https://orkg.org/papers/R1469778/R1469970（面板3）。该记录对应于 Ghazy 等人的综述中表5的第一行数据 [25]，以及文章《通过原子层沉积在低温硅基底上制备的多晶铒掺杂 Y_3镓$_5$氧$_{12}$化物纳米薄膜及其电致发光性能的探索》 [68]。

在实践中，传统的调查问卷表或综述表格会被拆分，使每个表格行 成为一项独立的贡献 （一种实验配置或已报道器件），而每个列 则成为一项明确定义的属性。图1和图2以 Ghazy等人 （2025年） [25]的稀土原子层沉积综述中的表5为例对此进行了说明。面板（1）重现了原始表格；面板（2）展示了对应的 ORKG 比较（https://orkg.org/comparisons/R1469991），其中九个表格行变为九项贡献，表头则显示为 ORKG 属性（灰色列）。
这些属性分别对应于原子层沉积实验人员经常解读的性能描述符：

- **主体材料基质** (P180031)：掺入 Er^{3+} 的晶态或非晶态材料体系（例如，$Y_3Ga_5O_{12}$、Ga_2O_3）。

- **退火温度** (P180032)：用于激活掺杂剂、控制缺陷或提高结晶度的沉积后热处理。

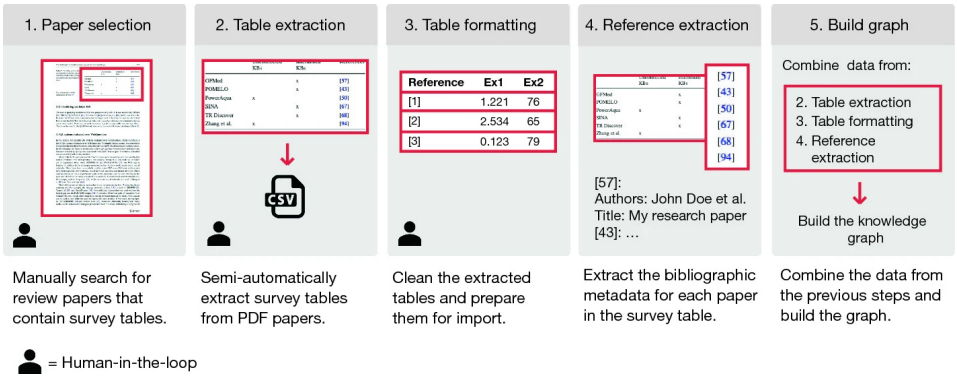


图3：将综述表格导入开放研究知识图谱（ORKG）的方法。图转载自 Oelen等人 (2020) 的《从综述文章表格创建学术知识图谱》 [45],，该技术首次发表于此文。

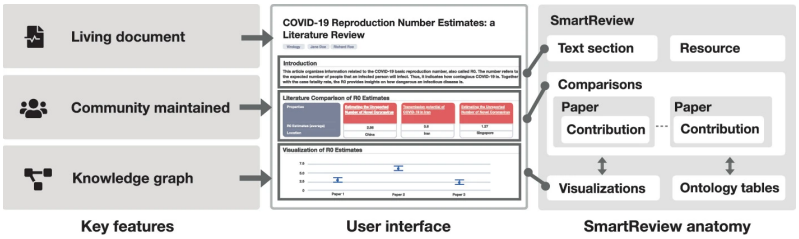


图4：SmartReviews 关键特征和结构的示意图。它们由若干构件组成，包括自然语言文本、比较和可视化。图转载自 Oelen等人 (2021) 的《SmartReviews：面向人类和机器可操作的综述文章表示》 [46], , SmartReview 方法首次发表于此文。

- **外量子效率（EQE）**(P180033)：注入的电荷载流子转化为发射光子的比例。
- **功率效率（PE）**(P180034)：单位电输入对应的光输出功率，此处报告的数值按 $\times 10^{-4}$缩放。
- **阈值电压**(P180035)：电致发光开始时的电压。
- **发射寿命**(P180036)：1530–1550nm Er^{3+} 发射的衰减时间，表明辐射效率和缺陷环境。
- **器件工作寿命（OLT）**(P180037)：MOSLED器件在运行期间维持可用输出的时长。

将这七个描述符编码为显式谓词后，单个 PDF 表格便转化为一个小型、可搜索的 Er^{3+}基 MOSLED 器件性能数据集。对于 ALD领域而言，ORKG 仍如同熟悉的对比表，但作为一种知识图谱表示，它还可以被搜索、筛选、链接到其他 ALD/E数据，并以静态 PDF 表格无法实现的方式通过程序重复使用。

对于 ORKG平台，Oelen 等人 (2020年) 首先引入了*ORKG 比较*这一专用内容类型，用于表示和比较研究贡献 [44]。从概念上讲，ORKG 比较是一种结构化对比表，其中：

- 每一列代表一项贡献，即特定研究论文中报道的一种器件或实验配置（例如 Ghazy 等人 (2025年) 的表5中列出的九种 Er^{3+}基 MOSLED 器件之一）。每项贡献都直接链接到其来源出版物的 ORKG 条目，从而确保完全可追溯，使底层论文能够被打开、查阅和核实；
- 每一行都是一个类型化属性：退火温度、外量子效率（EQE）、功率效率、阈值电压、发射寿命等数值量，以及

器件工作寿命（OLT）以带单位的数值形式存储，而主体材料基质等类别属性则以受控值存储，而非任意自由文本；

- 每个单元格都可以容纳类似资源的值，这意味着反复出现的实体（例如特定的主体材料基质，如 $Y_3Ga_5O_{12}$）只需作为可识别的资源表示一次（例如 https://orkg.org/resources/R1469904），便可在多个贡献和比较中重复使用；
- 该比较仍是一个持续演进的对象，可在新数据或更正信息可用时进行更新或扩展（例如，随着文献不断增加，纳入更多基于铕$^{3+}$的金属氧化物半导体发光二极管器件）；
- 反复出现的列集可以整理为模板，从而便于在多个综述表格中应用相同的属性模式（本例中包括：主体材料基质、退火温度、外量子效率、功率效率、阈值电压、发射寿命和 OLT）。

在一项后续研究中，Oelen 等人（2020年）展示了如何利用已发表的综述表格填充此类比较 [45]。如图3所示，该工作流程从 PDF 表格入手，提取单元格内容，将提取出的列与一个简单模式（即对各列含义的描述）对齐，然后由策展人检查并完善结果。已包含专门的“Ref.”或“Reference”列的表格尤其适用：每一行自然对应于一项与特定论文关联的贡献，而每个列标题都可作为候选 ORKG 属性。最终得到的是 *FAIR*（可发现、可访问、可互操作、可重用）科学数据 [62]：表格不再只是 PDF 中的图像，而是可供搜索、筛选并重新组合以开展新分析的机器可读记录。综述表格的情境提供了一种直接将行和列映射到 ORKG 比较的方法，但同一理念也适用于其他类型的科学知识（例如文本摘要、图组或实验方案描述）。在这些情况下，必须先确定或设计比较属性——即用于比较不同研究的各个方面——之后才能创建可重复使用的比较模板。

Oelen等（2021年）[46]在 SmartReviews 框架中进一步拓展了这些理念。SmartReview 并非一种新型比较，而是将传统综述文章与其底层 ORKG 比较及据此生成的可视化内容打包在一起的一种方式。换言之，SmartReview 将（i）普通文本、（ii）一个或多个 ORKG 比较，以及（iii）由结构化数据直接驱动的图表整合在一起。这样一来，综述既可以作为普通学术文章阅读，也可以作为经过整理的小型数据资源使用。对于 ALD 领域而言，这意味着稀土 MOSLED 性能表或 SiO_2 原子层刻蚀汇总表等表格不再彼此孤立：人们可以交互式地探索这些表格、将其用于绘制新的图表，或随着新论文的发表而更新内容。

在本文余下部分中，我们将采用这些 ORKG 概念，把关于金属氧化物和 SiO_2 的 ALD/E 综述表格编码为比较，为工艺参数和机理定义可重复使用的属性模板，并将它们整合成 SmartReview 风格的成果，以支持跨表格查询，并促进其在 ALD/E 领域的长期重复使用。

3 ORKG 中机器可处理的原子层沉积/原子层刻蚀综述表格与问答

在本节中，我们展示如何将原子层沉积/原子层刻蚀文献中的综述表格建模为开放研究知识图谱（ORKG）中的机器可处理的比较。表格一旦以这种结构化形式表示，就可以通过声明式查询检索、筛选和组合其中的条目，而无需手动浏览 PDF文档。为说明这一做法，我们将三篇近期综述论文中的七个原子层沉积表格，以及六篇综述论文中的十一个原子层刻蚀表格编码为 ORKG 比较。随后，我们围绕这些表格的内容提出一系列自然语言问题；针对每个问题，我们概述相应的查询，并展示完整或部分结果。我们的目标并非得出新的科学结论，而是证明，典型的原子层沉积/原子层刻蚀综述内容在转化为机器可处理的数据后，可以通过确定性查询进行检索，从而确保所检索数据的事实准确性和依据充分性。

原子层沉积(ALD)已得到广泛综述，从Suntola的早期奠基性文章 [59]，到随后关于前驱体和表面化学的基础性综述 [37,54]，再到van Ommen、Goulas和Puurunen撰写的综合性《柯克-奥思默百科全书》章节 [60] 以及George广受引用的领域综述 [22]。以工艺为重点的调查问卷

例如Miikkulainen 等 [41] 的研究，以及Popov 等 [52] 近期的更新，进一步梳理了ALD化学体系和温度窗口的发展，而Kessels 等在*Nature ReviewsMethodsPrimers*上发表的文章则概括了当前ALD应用的广度 [36]。

为展示具体原子层沉积场景中的机器可操作建模，我们选取了三篇近期综述论文。选取的主要标准是：每篇论文至少包含一个实质性的综述表格，用于比较多项研究中的参数或性能指标，并且包含专门的“Ref.”或“References”列。这与 ORKG 出版模式自然契合，因为表格行会链接回原始论文，如图2所示。相关论文的建模选取由本文第一作者决定。在以下小节中，我们将这些表格转换为 ORKG 比较，并使用两类示例问题对其进行查询：简单查询，即在单个表格上使用简单筛选条件进行查询；以及复杂查询，即组合多个属性、连接多个表格或计算导出量。

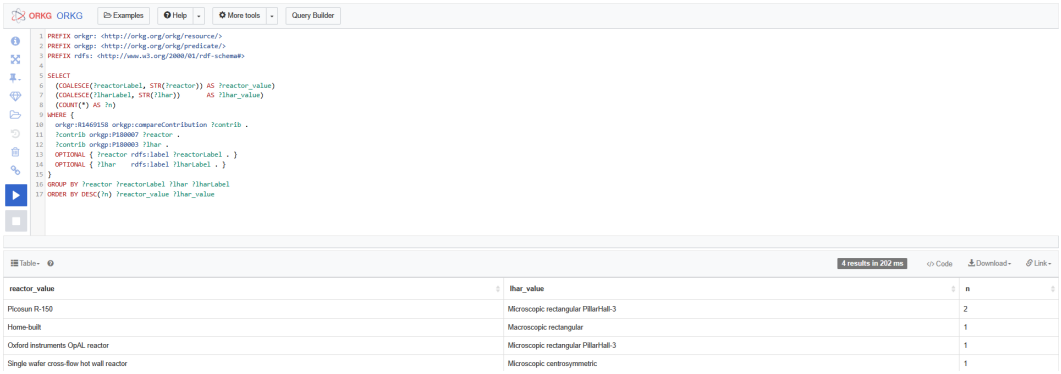


图5：ORKG平台提供的可视化 SPARQL 编辑器（https: //orkg.org/sparql/）截图，展示了对研究论文“基于饱和曲线的原子层沉积共形性分析：横向高深宽比通道中的氧化铝” [66], 中表2的 Q.1 的执行结果。该内容在 ORKG 中被建模为机器可操作的比较 R1469158。该查询检索了不同研究中反应器类型与横向高深宽比结构的共现情况，展示了不同实验设置如何将沉积系统与测试几何结构配对。可通过 https://tinyurl.com/lhr-reactor 重现该持久化查询网址。

3.1 论文——基于饱和曲线的横向高深宽比通道原子层沉积共形性分析

我们首先介绍 Yim等 [66],人的论文，其表2在 ORKG 中被建模为可机器处理的比较 https://orkg.org/comparisons/R1469158。该表报告了横向高深宽比（LHAR）结构中进行 Al_2O_3 原子层沉积时，三甲基铝（c_{TMA}）和 H_2O（c_{H2O}）的黏附系数，以及沉积温度、LHAR 几何结构、ALD循环次数和反应器配置等工艺条件。由于该表包含专门的“参考文献”列，将每一行与其来源文章关联起来，因此它很适合采用 ORKG 比较模型，其中每一行数据都可追溯至相应出版物。

在 ORKG 中，表2的比较 https://orkg.org/comparisons/R1469158 以可引用数据集的形式发布 [67] ，并拥有自己的 DOI：10.48366/R1469158。如图1的面板2所示，其结构与传统电子表格略有不同：每一列都是一个 ORKG 贡献（来自特定文章的一台设备或一种实验配置），每一行则对应一项属性（例如反应器类型、LHAR 几何结构、c_{TMA}），其数值可在不同贡献之间进行比较。换言之，原始综述表格被“横置”了，但所有数值和标签均得以保留，且每个数据点仍与其来源论文关联。

对于科学界而言，ORKG数字图书馆中的这种比较即数据集出版模式，相较于将静态CSV上传到 Zenodo 等存储库，提供了两项切实的激励。首先，经过精心整理的综述表格会成为拥有独立 DOI 的独立研究对象，可脱离文章单独引用，从而让其底层定量综合更易被关注，并可在后续工作、基准研究或建模工作中重复使用。其次，这项比较是一种动态资源：随着后续研究带来新的贡献，会添加为额外的列，而共享行集（属性）则保持稳定。更新版本可以重新发布

并采用新的数字对象标识符；ORKG 后端的版本控制系统会保留所有早期快照。从概念上讲，这类似于维基百科式的更新模式——随着新的原子层沉积结果出现，任何人都可以扩展这项比较——但还额外具备一项优势：每个已发布状态都会被冻结、加上时间戳并可供引用。以 LHAR 保形性研究等快速发展的领域为例，这使科学界能够维护最新且可由机器处理的比较，同时不丢失对早期时点知识状态的追踪。

表 1：针对 Yim 等人表 2 的问题 2 答案表， [66], 在 ORKG 中建模为比较 R1469158，列出了报告在 PillarHall-3 几何结构中 300 °C 下 c_{TMA} 数值的贡献。**C**

ontrib.	**研究问题**	**温度。[°C]**	**LHA R 类型**	**cTMA**
R1469105	原子层沉积中前驱体黏附系数的变化在高深宽比结构上	300	微观矩形 PillarHall-3	0.00572
R1469111	原子层沉积中前驱体黏附系数的变化在高深宽比结构上	300	微观矩形 PillarHall-3	0.00400

简单查询。 本文介绍的第一个查询是以下自然语言问题：“各项研究报告了哪些反应器类型与 *LHAR*结构的组合？每种组合出现的频率是多少？”将表2表示为 ORKG 比较后，就可以通过程序回答这类问题，而不必手动逐一查阅多份 PDF 表格。因此，我们将其实现为 SPARQL 查询——知识图谱的标准声明式查询语言，其作用类似于关系数据库中的 SQL——因为它能以透明且可复现的方式搜索、筛选和汇总结构化表格数据。在图5所示的具体查询中，顶部的 PREFIX声明为 ORKG资源（orkgr:）、属性（orkgp:）和标准 RDF术语（rdfs:）定义了简写名称；SELECT子句指定结果表中作为列显示的变量（此处为反应器类型、LHAR结构以及它们共同出现的次数）；WHERE块则描述了 ORKG 知识图谱中的链接模式，这些链接将每项比较贡献与其反应器及 LHAR 特征关联起来。OPTIONAL子句会在可用时提取人类可读的标签，而 GROUP BY 和 ORDER BY 则用于汇总和排序输出。这样，SPARQL 就能直接针对综述表格的机器可操作表示，以简洁易读的方式表达“列出所有反应器—几何结构组合及其出现频率”。

读者自然可能会问：“如何为自己的自然语言问题获取 SPARQL 查询？”有两种实用方式。第一，可以熟悉这些基本结构并手动编写 SPARQL；对于熟悉 SQL 的读者来说，上述类比提供了一个易于理解的起点，而 ORKG 可视化编辑器则通过并排展示查询及其结果，进一步降低了入门门槛。第二，现代大语言模型（LLM）在获得相关属性标识符 [64, 17]后，越来越擅长将自然语言问题转换为有效的 SPARQL 查询；尤其是，Avila等人 [8] 明确评估了 ChatGPT 在文本到SPARQL生成方面的表现。在实际操作中，用户可以用通俗的英语表述问题，请 ChatGPT 等大语言模型为感兴趣的 ORKG 比较拟定 SPARQL 查询，然后在 ORKG 编辑器中运行或调整该查询。这种工作流程大幅降低了与机器可操作的科学数据交互的门槛，同时保留了完全的透明度和可复现性。

我们第一个问题的 SPARQL 实现可通过 https://tinyurl.com/ lhr-reactor 重新运行。图5展示了这一查询在 ORKG平台的可视化 SPARQL 编辑器中的形式及其结果表，说明如何从表2的机器可操作表示中直接提取所有反应器与 LHAR 的共现信息。

复杂查询。 本研究介绍的第二个查询，也是第一个复杂查询，是以下这个带约束的问题：“在 *PillarHall-3* 结构中，温度为 *300 °C* 时，各项研究报告了哪些 c_{TMA} 数值？”在本文中，当查询超出简单的单属性筛选——例如在同一条件集中组合多个属性、关联多个比较，或计算导出量——我们便将其称为复杂查询。在这个首个示例中，查询仍针对单个比较，但同时限定了 ORKG 知识图谱中的三个显式属性：温度（P37561）、LHAR 类型（P180003）和 TMA 黏附系数 c_{TMA}（P180008）；此外，还可选地通过 P32 记录相关的研究问题。我们将此问题实现为 SPARQL 查询，可通过 https://tinyurl.com/pillarhall3-ctma 重新运行

针对同一个 ORKG 比较 https://orkg.org/comparisons/R1469158。结果汇总于表 1，以完全可由机器查询的形式重现了原始表格中的相关行。

总的来说，这两个问题表明，关于 LHAR 保形性的传统原子层沉积综述表格可以转化为结构化 ORKG 比较，并使用 SPARQL 进行确定性查询，而无需更改底层实验数据。我们强调，这些查询是方法论上的示范，并非对任何特定测试结构、反应器技术或商业实现的认可。如上所述，底层综述论文由第一作者通过独立文献检索选定，检索依据是预先制定的纳入标准（包含实质性的多行综述表格，且设有专门的“Ref.”/“References”列，并具有足够的引用影响力）；在此之前，这些论文的任何作者均未受邀担任本文的共同作者。因此，原综述的作者并未影响示例表格的选择。

3.2 论文——1988 至 2023 年颗粒材料的原子层沉积：技术、材料与应用的定量综述

作为我们选取的第二篇原子层沉积综述论文，我们选择了 Piechulla等人(2025年) [51]撰写的综合性定量综述。该文综合了三十多年来原子层沉积应用于颗粒材料的研究，对 799 篇出版物进行了系统分析，以梳理反应器技术、涂层材料、反应物、载体和工艺条件。该研究采用的定量方法与我们在 ORKG 中呈现结构化科学知识的目标十分契合，因为它提取了反复出现的实验参数，并在大量文献中对这些参数进行了交叉引用。
本文中，我们对表 3 和表 4 进行建模。这两张表都设有专门的参考文献 列，将报告的数据与来源出版物关联起来，因此特别适合用于 ORKG 比较建模，以及可追溯、可查询的数据集成。

3.2.1 表3的机器可操作建模

Piechulla等人 （2025年） [51] 的综述中的表3总结了荧光粉材料的原子层沉积涂层研究。该表列出了每项研究的荧光粉基质（例如 $BaMgAl_{10}O_{17}:Eu^{2+}$、$Sr[LiAl_3N_4]:Eu^{2+}$、$Zn_2SiO_4:Mn^{2+}$）、原子层沉积涂层材料（例如 SiO_2, Al_2O_3, TiO_2, MgO、氧化锌、氮化钛）、前驱体方案（例如 TEOS/H_2O、三甲基铝/H_2O、$TiCl_4/H_2O$、DEZ/H_2O）、涂层厚度（通常为 2–50 nm，或以循环次数表示）以及相应的参考文献。

我们将此表建模为 ORKG 比较 “荧光粉材料上原子层沉积涂层的比较研究”，网址为 https://orkg.org/comparisons/R1469383 [48]。在这种机器可操作的形式中，每条记录都与其来源出版物相关联，并且所有字段均表示为显式属性，因此可以直接通过 SPARQL 查询回答结构化问题，例如“哪些荧光粉采用了硅涂层？”或“对于 Eu^{2+} 激活基质上的 Al_2O_3 涂层，使用了哪些前驱体方案？”，而无需重新阅读底层 PDF 表格。

Tab表 2：针对 Piechulla等人表3的 Q.3 的答案表， [51], 在 ORKG 比较中建模为 s Co较 R1469383。

贡献	荧光粉	涂层
R1469163	BaMg铝$_{10}O_{17}$:铕$^{2+}$	SiO_2
R1469173	BaMg铝$_{10}O_{17}$:铕$^{2+}$	SiO_2
R1469216	Sr[锂铝$_3N_4$]:铕$^{2+}$	SiO_2

简单查询。 在我们的问答数据集中，第三个问题是：“哪些荧光粉涂覆了二氧化硅$_2$？”该问题基于 ORKG 的表3比较提出。将表格建模到 ORKG 后，便可通过程序回答这一问题，而无需重新阅读原始综述。我们将其实现为一个 SPARQL 查询，可通过 https://tinyurl.com/Phosphor-SiO2-ALD 重新运行。不同于返回自由文本回答的传统问答任务，我们的系统会生成一份综合性的、可追溯证据的答案表：表2列出了比较中报告的所有采用二氧化硅$_2$ 原子层沉积涂层的荧光粉材料，并将每种材料明确关联到其原始贡献。这种表格化输出减少了冗长且可能重复的叙述，并支持对底层数据进行透明验证。

表3：针对 Piechulla等人表3中问题4的答案表 [51], , 在 ORKG 中建模为比较 R1469383。

贡献	荧光粉	涂层	前驱体	temp (°C)	厚度 (nm)
R1469207	Sr_2硅$_5N_8$:铕$^{2+}$	Al_2O_3	TMA/H_2O	30	8
R1469207	Sr_2硅$_5N_8$:铕$^{2+}$	Al_2O_3	TMA/O 3	30	8

复杂查询。 第四个问题是：“在 Eu^{2+}掺杂的红色荧光粉中，哪些在 $\leq$ 150 ℃ 下沉积的原子层沉积涂层达到了 $\leq$ 20 nm 的最佳厚度，并使用了哪些前驱体？”该自然语言问题的 SPARQL 查询实现可通过 https://tinyurl.com/Red-Eu2-ALD-thinlowT 重新运行。如前文所述，本文中的答案是基于原始问题整理并综合得出的表格化结果；表3列出了满足这些查询条件的所有 Eu^{2+}掺杂红色荧光粉、涂层、前驱体、沉积温度和最佳厚度。

3.2.2 表4的机器可操作建模

下一个 ORKG 比较（网址：https://orkg.org/comparisons/R1469594 [49], ）仿照 Piechulla 等人（2025年） [51]的表4建模。该表汇总了药学相关颗粒载体上原子层沉积涂层的已报告厚度范围和每周期生长量（GPC）值。其中包含来自十篇出版物的25条实验条目，涵盖多种药物载体和生物分子基底（例如乳糖、布地奈德、吲哚美辛、HPV衣壳蛋白亚单位和蛋白质），以及原子层沉积工艺参数，如前驱体化学、沉积温度、涂层厚度和GPC值。在 ORKG 中，这些信息以机器可操作的比较形式呈现，为本节中的问答示例提供依据。

表4：Piechulla等人表4中问题5的答案表， [51], 在ORKG中建模为 ComparisonR1469594。为简洁起见，载体、前驱体组合和温度均相同的行合并显示。

载体	前驱体	temp (°C)	厚度（nm）	贡献
布地奈德	硅Cl_4/H_2O	40	1.5–10.0	R1469559 , R1469560 , R1469558
布地奈德	TMA/H_2O	40	0.8–2.0	R1469545
布地奈德	TMA/O 3	40	20.0–50.0	R1469551 , R1469548 , R1469554
布地奈德	钛Cl_4/H_2O	40	3.0–15.0	R1469557 , R1469555 , R1469556
乳糖	三甲基铝/H_2O	30	2.5–11.0	R1469543 , R1469540
乳糖	TMA/O 3	30	3.8–10.3	R1469544
肌红蛋白	三甲基铝/H_2O	30	8.3	R1469580
泊沙康唑/羟丙甲纤维素醋酸琥珀酸酯	三甲基铝/H_2O	35	8.0–40.0	R1469589
泊沙康唑/羟丙甲纤维素醋酸琥珀酸酯	三甲基铝/H_2O	35	12	R1469588

简单查询。 第五个问题是：“在 $\leq$ 40 ℃ 下涂覆了哪些支撑材料，使用了哪些前驱体对，以及报告的涂层厚度是多少？”此问题的 SPARQL 查询实现可通过 https://tinyurl.com/less-than-40 重新运行。表格答案见表4，其中列出了所有满足 $\leq 40^\circ$C 条件的药学相关支撑材料、前驱体对、沉积温度和厚度范围。

表5：针对 Piechulla等人的表4中问题6的答案表 [51], , 在 ORKG 中建模为比较 R1469594。具有相同载体–前驱体–温度组合的行合并显示，以节省空间。

载体	前驱体	温度 (°C)	厚度 (nm)	每循环生长量 (nm)	类别	贡献
乳糖	三甲基铝/H_2O	30	2.5	0.47	平均值	R1469540, R1469543
乳糖	TMA/O_3	30	3.8	0.47	平均值	R1469544
布地奈德	TMA/H_2O	40	0.8	0.20	slow	R1469545
布地奈德	$TiCl_4$/H_2O	40	3.0	0.30	n/a	R1469555 , R1469556 , R1469557
布地奈德	硅Cl_4/H_2O	40	1.5	0.10	n/a	R1469558 , R1469559 , R1469560
肌红蛋白	三甲基铝/H_2O	30	8.3	0.21	slow	R1469580
泊沙康唑/羟丙甲纤维素醋酸琥珀酸酯	三甲基铝/H_2O	35	12	0.40	平均值	R1469588
泊沙康唑/羟丙甲纤维素醋酸琥珀酸酯	三甲基铝/H_2O	35	8.0	0.40	平均值	R1469589

复杂查询。 第六个问题是：“在低温运行（< 70 °C）中，哪些运行制得了薄涂层（< 20 *nm*）？这些运行报告了哪些载体材料、前驱体对、温度、厚度和每循环生长量（*GPC*）值？对于氧化铝涂层，这些*GPC*值如何分类为慢、平均或快？”可通过以下链接重新运行 `https://tinyurl.com/pharma-hard-query`SPARQL 查询。表格形式的答案见表5，其中列出了满足查询条件的所有载体材料、前驱体对、温度、厚度、GPC值及相应的GPC类别。

3.3 论文——基于稀土元素的功能薄膜原子层沉积和分子层沉积

现在，我们来考察 Ghazy等人 (2025年) [25], 题为《基于稀土元素的功能薄膜原子层沉积和分子层沉积》的综述。与第一篇聚焦于高深宽比测试结构中的保形性、第二篇对颗粒型原子层沉积系统进行大规模定量综合的论文不同，这篇综述着眼于稀土基原子层沉积和分子层沉积薄膜的化学性质与功能属性。我们将该论文中的表2、表3、表4和表5建模为机器可操作的 ORKG 比较，每个比较均配有简单查询和复杂查询。结合前两篇综述论文，这些模型涵盖了广泛的原子层沉积相关信息，并展示了如何在 ORKG 中以统一、可查询的知识图谱框架表示多样化的实验和材料数据。

3.3.1 表格的机器可操作建模 le 2

我们首先将 Ghazy 等人（2025年）的表2转换为 ORKG 比较，可通过 https: //orkg.org/comparisons/R1469955 [1]访问。这样一来，它便可供引用、重复使用和版本管理，成为一份持续更新的综述表格，并具有完整的溯源信息，且可访问早期版本。表2的标题为“不同稀土物种在原子层沉积中作为掺杂剂的应用”，列出了原子层沉积工艺中使用的主体材料和稀土掺杂剂。表中涵盖氧化物（例如 Al_2O_3, TiO_2, HfO_2）、硫化物（例如 ZnS、SrS、CaS）和石榴石（例如 $Y_3Ga_5O_{12}$,$Lu_3Al_5O_{12}$），并列出了已报告的应用，如发光、磁学、栅介质、电解质和测温。每条记录都将特定的基质材料—掺杂剂组合与一条或多条文献参考关联起来，为下文讨论的简单查询和复杂查询提供了结构化基础。

简单查询。 第七个问题是：“对于每种应用（发光、存储器、栅介质、电解质等），哪些稀土掺杂剂出现于数量最多的不同基质中？”此问题的 SPARQL 查询可通过 https: //tinyurl.com/versatile-dopants，在综述文章表2的 ORKG 比较R1469955中重新运行。汇总答案见表6，其中按应用列出了稀土掺杂剂及其对应的基质数量。

表6：针对 Ghazy 等人表2的 问题7 答案表 [25], ，该表在 ORKG 中被建模为比较 R1469955。对于每种应用，该表列出了稀土掺杂剂及不同主体材料的数量。

应用 掺杂剂	主体材料
电解质 Y、Dy、La	2、1、1
栅极二- 镧、镝、铒、钆、镨、钇	2, 1, 1, 1, 1, 1
电气	1
界面相 镝、镧、钇	1, 1, 1
发光铒、铕、铽、钕、铈、镝、钬、镧、镨、钐、铥、镱、（铒、镱）	12, 5, 5, 3, 2, 2, 2, 2, 2, 2, 2, 1
磁学 镝、铒、铕、钬、镧、钕、镨、钐、Tb, Tm, Yb	各 1 个
存储器 钆、镧、钇、镝、铒、镨	2、2、2、1、1、1
光电子学镧	1
测温镝、铒、铕、钬、钕、钐、铽、Tm, Yb	各1个
波导（铒、镱）、铒、镱	1 个

表7：Ghazy等人关于问题8的表2的答案表 [25], 在 ORKG 中建模为比较R1469955。对于至少出现在两个应用领域中的每种基质材料，表中列出了其应用和稀土掺杂剂。

host	应用	掺杂剂
铝$_2$O$_3$	发光, 波导	(铒,镱), 铒, 镱
HfO$_2$	Gate 电介质, 存储器	镧, 镝, 铒, 钆, 镨, 钇
TiO$_2$	发光, Mag- 磁学	铒、铕、镝、钬、镧、钕、镨、钐、铽, Tm, Yb
ZrO$_2$	电解质, 相间- 相	镝、镧、钇

复杂查询。 第八个问题是：“哪些主体材料出现在两个或更多不同的应用领域中，它们分别对应哪些掺杂剂？”此问题的 SPARQL 查询实现可通过 https://tinyurl.com/cross-functional-hosts，在综述文章表2的 ORKG 比较R1469955中重新运行。汇总答案见表7，其中列出了所有至少出现在两个应用领域中的主体材料，以及相应的应用和稀土掺杂剂。

3.3.2 表3的机器可操作建模

我们将 Ghazy 等人（2025年）的表3建模为 ORKG 比较，网址为 https://orkg.org/comparisons/R1471077 [2]。该比较记录了每种已报道的稀土（R^{3+}）原子层沉积工艺的沉积材料、金属前驱体、共反应物、每循环生长量（GPC）、沉积温度范围和文献参考，从而对254种进行比较的原子层沉积工艺形成了结构化表征。该表涵盖了广泛的二元和三元体系，包括稀土与 Al、Ti、Ni、Ga、W、Co、Mn、Fe、Hf、Y、Sc、Lu 和 Li 的组合，其中许多采用超循环策略来控制元素比例。在 ORKG 中，这些信息构成了机器可操作的基础，可用于下文介绍的稀土原子层沉积化学简单查询和复杂查询。

简单查询。 第九个问题是：“列出所有在低沉积温度（$\leq 250\ ^\circ C$）下实现高每周期生长量（*GPC* ≥ 1 *Å*）的稀土原子层沉积工艺。报告材料、金属前驱体类别、共反应物、*GPC* 和温度；按 *GPC* 排序。”此问题的 SPARQL 查询实现可通过以下链接重新运行：`https://tinyurl.com/ald-highgpc-lowtemp` 该链接对应于综述文章表3中的 ORKG 比较R1471077。表格答案见表8，其中

表8：Ghazy等人表3中问题9的答案表 [25], , 在ORKG中建模为比较R1471077。对于每种配体家族，该表列出了高每循环生长量（≥1 Å）和低温度（≤ 250 °C）稀土原子层沉积工艺的代表性材料、每循环生长量和温度范围，以及共反应物。

配体家族	代表性材料	每循环生长量 [Å]	温度 [°C]	共反应物
Cp衍生	钇$_2O_3$, 铒$_2O_3$, 钆$_2O_3$, 铽$_2O_3$, GdAlO $_3$	1.0–4.5	175–250	H_2O, O_3, TMA
脒基 / 甲脒基-nate	镧$_2O_3$, 铈O_2, 镨$_2O$ $_3$, Gd_2O_3	1.1–2.8	200–240	H_2O, O $_3$
胍基化物	铈O_2, 钇$_2O_3$, 镝$_2O_3$,钆$_2O_3$	1.0–2.1	150–200	H_2O
β-二酮	镧F_3, 钇F_3, 钬F_3	1.0–2.0	200–225	TiF_4, NbF_5

按配体家族对筛选出的稀土原子层沉积工艺进行分组，并概述代表性材料、GPC范围、温度范围和共反应物。

表9：针对 Ghazy 等人表3提出的问题10的答案表，已在 ORKG 中建模为比较 R1471077。对于 200–300 °C 温度区间内的 Y_2O_3 工艺，该表报告了 ALD 模式、平均 GPC、记录计数和示例共反应物。

Mode	平均每循环生长量 [Å]	记录	示例共反应物
热原子层沉积	1.95	28	钇$(thd)_3$
等离子体增强原子层沉积	1.65	3	O_2 等离子体

复杂查询。 第十个问题是：“仅针对 *Y_2O_3* ，比较 *200–300* °C 温度范围内等离子体增强原子层沉积（*PE-ALD*）与热原子层沉积的平均每周期生长量（*GPC*）。返回模式（*PE* 与热）、平均*GPC*、记录数量以及共反应物示例列表。”此问题的 SPARQL 查询实现可通过 https://tinyurl.com/ald-y2o3-gpc200 重新运行，查询对象为综述文章中表3的 ORKG 比较R1471077。表9列出了汇总答案，包括该温度范围内热原子层沉积和 PE-ALD Y_2O_3 工艺的平均GPC、记录计数及共反应物示例。

3.3.3 机器可操作建模：表4

我们将 Ghazy 等人 （2025年）的表4建模为 ORKG 比较，网址为 https://orkg. org/comparisons/R1470110 [3]。该比较记录了每种已报告的 ALD/MLD工艺的稀土-有机杂化材料、金属前驱体（例如 $R(thd)_3$, $R(dpdmg)_3$）、有机连接体（例如 TPA、PDC、NDC）、每周期生长量（GPC）、沉积温度和文献参考。

因此，该表既涵盖了采用芳香族连接体的经典 $R(thd)_3$ 路线，也包括报告了更高每循环生长量的新型胍基化物化学体系。在 ORKG 中，这种结构化表示为下文介绍的稀土-有机杂化薄膜简易查询和复杂查询奠定了基础。

表10：针对 Ghazy 等人表 4 的问题11答案表 [25], , 在 ORKG 中建模为比较 R1470110。该表列出了每循环生长量为 ≥5 Å、沉积温度为 ≤250 °C 的稀土原子层沉积/分子层沉积杂化薄膜，包括材料体系、金属前驱体、有机前驱体、每循环生长量和温度。

材料体系	金属前驱体	有机前驱体	每循环生长量 [Å]	T [°C]
Ce–TPA	$Ce(DPDMG)_3$	TPA	5.4	200
Er–3,5PDC	$Er(DPDMG)_3$	3,5PDC	6.4	245
Eu–HQA	铕$(thd)_3$	HQA	7.3	210

简单查询。 第十一个问题是：“列出所有在低沉积温度（≤ 250 °C）下实现高每周期生长量（*GPC* ≥*5 Å*）的稀土原子层沉积/分子层沉积杂化薄膜。报告材料体系、金属前驱体家族、有机前驱体、每周期生长量和温度；按每周期生长量排序。” 此问题的 SPARQL 查询实现可通过 `https://tinyurl.com/HiGLoT-query` 重新运行

针对综述文章中表4的 ORKG 比较 R1470110。表格中的答案见表10，其中列出了相应的高生长速率、低温杂化薄膜工艺及其沉积条件。

复杂查询。 第十二个问题是：“按有机连接体族（对苯二甲酸酯、吡啶二甲酸酯、萘二甲酸酯、吡嗪基、其他）对稀土原子层沉积/分子层沉积杂化薄膜进行分组，并计算每组的平均每周期生长量（*GPC*），仅考虑在 $\leq 250\,^{\circ}\mathrm{C}$ 温度下沉积的薄膜。报告连接体族、平均*GPC*和薄膜数量；按平均*GPC*排序。”此问题的 SPARQL 查询实现可通过 `https://tinyurl.com/aldmld-linker-gpc`在综述文章表4的 ORKG 比较 R1470110 上重新运行。

3.3.4 表5的机器可操作建模

我们将 Ghazy 等人（2025年）的表5建模为一项 ORKG 比较，网址为 https: // orkg.org/comparisons/R1469991 [24], , 并使用模板 https://orkg.org/ templates/R1469957 进行实例化。该比较记录了每种 Er^{3+} 掺杂 MOSLED器件的基质、退火温度、外量子效率（EQE）、功率效率（PE）、阈值电压、发射寿命（τ）和器件工作寿命（OLT）。这种机器可操作表示为查询 Er^{3+} 掺杂 MOSLED 的基质组成和加工条件与器件性能之间的关系提供了基础。

表11：针对 Ghazy等人表5的 Q.13 答案表 [25], , 在 ORKG 中建模为比较 R1469991。该表列出了 EQE 为 $\geq$10%且报告了阈值电压的 Er^{3+} 基 MOSLED基质，以及 EQE、电压、退火温度、发射寿命（τ）和工作寿命（OLT）。

基质	外量子效率 [%]	电压 [V]	退火 [°C]	寿命 τ [ms]	工作寿命（OLT）[小时]
铝$_2$氧$_3$:镱	24	14	–	0.9	–
Ga_2O_3	36	15	900	2.04	100
镥$_3$铝$_5$氧$_{12}$	10	50	1100	1.65	–

简单查询。 第十三个问题是：“在比较中的铒$^{3+}$ 金属氧化物半导体发光二极管中，哪些主体基质在最低阈值电压下实现了较高的外量子效率（外量子效率 $\geq$*10%*），报告的退火温度和寿命（τ、工作寿命（*OLT*））分别是多少？”该问题的 SPARQL 查询实现可通过 https://tinyurl.com/mosled-high-eqe 在综述文章表 5 的 ORKG 比较 R1469991 上重新运行。表11列出了汇总答案，其中包含相应的主体基质以及外量子效率、阈值电压、退火温度、发射寿命和工作寿命。

表12：Ghazy等人的表 5 中问题14的答案表，该内容在 ORKG 中建模为比较 R1469991。对于每种基于铒的金属氧化物半导体发光二极管主体基质，表中列出了外量子效率、工作电压、衍生的外量子效率/电压指标、退火温度、发射寿命（τ）和工作寿命（OLT）。

基质	外量子效率 [%]	电压 [V]	外量子效率/电压	年退火温度 [°C]	τ [毫秒]	OLT [h]
Ga_2O_3	36.0	15	2.40	900	2.04	100
铝$_2$氧$_3$:镱	24.0	14	1.71	–	0.90	–
镱$_3$铝$_5$O$_{12}$	5.2	5	1.04	1100	1.20	–
镥$_3$铝$_5$O$_{12}$	10.0	50	0.20	1100	1.65	–
Yb_2O_3	8.5	60	0.14	1000	–	–
钇$_3$镓$_5$O$_{12}$	2.5	25	0.10	800	2.26	49
钆$_3$镓$_5$O$_{12}$	1.9	30	0.06	1000	1.59	–
$ZnGa_2O_4$	0.3	30	0.01	800	0.82	5.5

复杂查询。 第十四个问题是：“计算每伏特效率指标（外量子效率/电压），以近似衡量外量子效率–电压权衡，并据此对主体基质进行排序。”SPARQL 查询

此问题的实现可通过综述文章中表5的 ORKG 比较 R1469991 在 https://tinyurl.com/mosled-pareto-score 重新运行。表格答案见表12，其中列出了每种基质的外量子效率、工作电压、外量子效率/电压、退火温度、发射寿命和工作寿命。

3.3.5 跨建模表格查询

综述表格的机器可操作建模不仅支持表内查询，还支持跨表查询，从而整合互补信息，形成统一的分析视图。这使用户能够得出汇总且精确的见解，而这些见解无法仅从任何单张表格中获得。本节将展示如何联合查询 Ghazy等人 (2025年) 的表2、表3和表5：表2列出了铒掺杂或具有发光特性的基质材料，表3提供了相应的原子层沉积工艺配方，表5则报告了掺杂铒体系的金属氧化物半导体发光二极管器件指标。表4聚焦于稀土-有机杂化薄膜，与表5中的MOSLED数据重合有限，因此将其视为可选内容，不纳入以下示例。

表13：问题15的答案表，合并了 Ghazy等人的表 3 和表 5。 [25], 在 ORKG 中建模为比较 R1471077（表 3）和 R1469991（表 5）。对于每种基于铒$^{3+}$的金属氧化物半导体发光二极管基质，表中列出了所报告沉积温度最低的相应原子层沉积工艺，包括工艺材料、金属前驱体、共反应物、前驱体类别、每循环生长量和沉积温度，以及外量子效率。

基质	EQE [%]	工艺材-rial	金属前驱体-sor	Co-反应物	族	GPC [埃/周期]	T [°C]
钆$_3$镓$_5$氧$_{12}$	1.9	钆$_3$镓$_5$氧$_{12}$	钆(thd)$_3$	TEGa	R(thd)$_x$	0.19	350
镥$_3$铝$_5$氧$_{12}$	10.0	镥$_3$铝$_5$氧$_{12}$	镥(thd)$_3$	TMA	R(thd)$_x$	0.19	350
钇$_3$镓$_5$O$_{12}$	2.5	钇$_3$镓$_5$O$_{12}$	钇(thd)$_3$	Et$_3$镓	R(thd)$_x$	0.23	350
镱$_2$O$_3$	8.5	镱$_2$O$_3$	镱(MeCp)$_3$	H$_2$O	其他	1.00	180
镱$_3$铝$_5$O$_{12}$	5.2	镱$_3$铝$_5$O$_{12}$	Yb(thd)$_3$	TMA	R(thd)$_x$	0.20	350

简单查询 1. 第十五个问题是：“展示 *Er-MOSLED* 基质的原子层沉积配方及其器件性能。”此问题的 SPARQL 查询实现可通过 https: //tinyurl.com/t3t5-lowT 在 ORKG 比较 R1471077（表 3）和 R1469991（表5）上重新运行。对于每种基于 Er^{3+} 的 MOSLED 基质，该查询会提取表5中报告的最高外量子效率，并将其与表3中对应的 ALD 工艺参数（工艺材料、金属前驱体、共反应物、前驱体类别、每循环生长量和沉积温度）关联起来。汇总答案见表13。

表14：Q.16 的答案表，关联 Ghazy等人表2和表5中的数据，这两张表在 ORKG 中建模为比较R1469955（表 2）和比较R1469991（表5）。对于同时出现在两张表中的每种发光稀土掺杂材料，该表列出了外量子效率、阈值电压、功率效率（PE）、退火温度、发射寿命（τ）和工作寿命（OLT）。

材料	外量子效率 [%]	阈值 [V]	PE [$\times 10^{-4}$]	Ann. T [°C]	τ [毫秒]	OLT [小时]
Ga$_2$O$_3$	36	∼15	81	900	2.04	100
钆$_3$镓$_5$氧$_{12}$	1.9	30	3.6	1000	1.59	–
镥$_3$铝$_5$氧$_{12}$	10	50	12	1100	1.65	–
钇$_3$镓$_5$氧$_{12}$	2.5	25	3.8	800	2.26	49
Yb$_2$O$_3$	8.5	60	10	1000	–	–
镱$_3$铝$_5$氧$_{12}$	5.2	∼5	4.8	1100	1.20	–
锌镓$_2$氧$_4$	0.3	30	–	800	0.82	5.5

简易查询 2。第十六个问题是：“*ALD* 掺杂剂概述中列为掺杂体系的哪些发光材料，也在稀土 *MOSLED* 性能比较中报告了实测 *MOSLED* 性能？”此问题的 SPARQL 查询实现可通过 https://tinyurl.com/LumMOSLED-query 在 ORKG 比较R1469955（表2）和 R1469991（表5）上重新运行。对于同时出现在两张表中的每种发光主体材料，查询会提取其实测性能参数——外量子效率、阈值电压、功率效率（PE）、

退火温度、发射寿命（τ）和工作寿命（OLT）。表格化的答案见表14。

表15：问题17的答案表，合并了 Ghazy 等人的表3和表5，并在 ORKG 中建模为比较R1471077（表3）和 R1469991（表5）。对于每种发光MOSLED主体基质，该表列出了外量子效率以及与之对应的 ALD 工艺参数：工艺材料、金属前驱体、共反应物、每循环生长量和沉积温度。

基质	工艺材-rial	EQE [%]	金属前驱体	Co-反应物	GPC [埃/周期]	T [°C]
镥$_3$铝$_5$O$_{12}$	镥$_3$铝$_5$O$_{12}$	10.0	镥(thd)$_3$	TMA	0.19	350
镱$_2$O$_3$	镱$_2$O$_3$	8.5	镱(thd)$_3$	O_3	0.15–0.20	250–400
镱$_2$O$_3$	镱$_2$O$_3$	8.5	镱(thd)$_3$	O_3	0.15–0.20	250–400
镱$_2$O$_3$	镱$_2$O$_3$	8.5	镱(thd)$_3$	O_3	0.15–0.20	250–400
镱$_2$O$_3$	镱$_2$O$_3$	8.5	镱(MeCp)$_3$	H_2O	1.00	180–375
镱$_3$铝$_5$O$_{12}$	镱$_3$铝$_5$氧$_{12}$	5.2	钇镱(thd)$_3$	TMA	0.20	350
钇$_3$镓$_5$氧$_{12}$	钇$_3$镓$_5$氧$_{12}$	2.5	钇(thd)$_3$	Et$_3$镓	0.23	350
钆$_3$镓$_5$O$_{12}$	钆$_3$镓$_5$O$_{12}$	1.9	Gd(thd)$_3$	TEGa	0.19	350

复杂查询 1. 第十七个问题是：“对于稀土 *MOSLED* 性能比较中列出的每种发光 *MOSLED* 主体材料，检索其在稀土 *ALD* 工艺概览中对应的 *ALD* 工艺参数——金属前驱体、共反应物、单位循环生长速率和沉积温度，并将这些参数与外量子效率（*EQE*）一并报告。”此问题的 SPARQL 查询实现可通过以下链接重新运行： `https://tinyurl.com/t3t5-complex` 检索范围为 ORKG 比较R1471077（表3）和 R1469991（表 5）。表15列出了表格形式的答案。

表16：Ghazy等人第 18 题的答案表，合并了表 3 和表 5，并在 ORKG 中建模为比较 R1471077（表3）和 R1469991（表 5）。对于两张表中均出现的每种基质，列出了合成温度、退火温度、外量子效率（EQE）、功率效率（PE）和发射寿命。

基质	合成温度 [°C]	退火温度 [°C]	外量子效率 [%]	功率效率 [%]	寿命 [ms]
Ga_2O_3	350	900	36.0	81.0	2.04
铝$_2$氧$_3$:镱	350		24.0	28.0	0.90
镥$_3$铝$_5$氧$_{12}$	350	1100	10.0	12.0	1.65
铝$_2$氧$_3$ – 钇$_2$氧$_3$	350	800	8.7	12.0	
Yb_2O_3	350	1000	8.5	10.0	
镱$_3$铝$_5$氧$_{12}$	350	1100	5.2	4.8	1.20
钇$_3$镓$_5$氧$_{12}$	350	800	2.5	3.8	2.26
钆$_3$镓$_5$氧$_{12}$	350	1000	1.9	3.6	1.59

复杂查询 2. 第十八个问题是：“表 *3* 和表 *5* 中，哪些稀土氧化物基质同时包含工艺层面的合成信息和器件层面的性能数据？合成和退火条件与其光学效率及寿命有何关系？”此问题的 SPARQL 查询可通过 https://tinyurl.com/t3t5-correlation 在 ORKG 比较 R1471077（表 3）和 R1469991（表 5）上重新运行。对于同时出现在这两个表格中的每种基质，查询会提取其合成温度、退火温度、外量子效率（EQE）、功率效率（PE）和发射寿命。汇总答案见表 16。

复杂查询 3. 第十九个问题是：“从表*2*中选出掺铒并用于发光的氧化物材料。结合它们的合成温度（表*3*）和外量子效率（表*5*），然后计算效率指数 =（外量子效率 / 合成温度 $\times$ *100*），以找出能在较低制备温度下实现高效率的材料。按此指数对材料排序。”此问题的 SPARQL 查询可通过`https://tinyurl.com/t2t3t5-complex` 在 ORKG 比较 R1469955（表2）、R1471077（表3）和 R1469991（表5）上重新运行。汇总答案见表 17。

作为本小节的最后一点，我们发布了从 Ghazy 等人（2025年）综述文章 《原子与分子层表格中整理出的全部四项机器可操作的 ORKG 比较

表 17：问题19的答案表，合并了Ghazy等人的表2、表3和表5，这些表在ORKG中分别建模为比较R1469955（表2）、R1471077（表3）和R1469991（表5）。对于出现在全部三个表格中的铒掺杂发光氧化物，列出了合成温度、外量子效率以及推导效率指数（外量子效率/合成温度 × 100）。

材料	合成温度 [°C]	外量子效率 [%]	效率指数
Ga_2O_3	250	36.0	14.4
Al_2O_3	250	24.0	9.6
Y_2O_3	140	8.7	6.21
Yb_2O_3	180	8.5	4.72
镥$_3$铝$_5$氧$_{12}$	230	10.0	4.35
镱$_3$铝$_5$O$_{12}$	350	5.2	1.49
钇$_3$镓$_5$O$_{12}$	350	2.5	0.71
钆$_3$镓$_5$O$_{12}$	350	1.9	0.54

基于稀土元素的功能薄膜沉积》 [25]：表2见 https://orkg.org/comparisons/R1469955，表3见 https://orkg.org/comparisons/R1471077，表4见 https://orkg.org/comparisons/R1470110，表5见 https://orkg.org/comparisons/R1469991。这四项比较共同整理于 ORKG 中，构成一篇采用智能综述格式的迷你综述文章（https://orkg.org/reviews/R1471098）。智能综述是原综述的精简、机器可操作版本，展示了结构化、可查询的原子层沉积数据表示如何补充并拓展传统学术出版。

至此，我们结束关于 ORKG 中 ALD 机器可处理比较的章节。总计有三篇近期 ALD 综述论文中的七个表格，从原始 PDF 格式转换为可引用、机器可处理的 ORKG 比较。基于这些表示形式，我们通过执行 19 条 SPARQL 查询，展示了机器可处理数据的优势。

接下来，我们转向原子层刻蚀（ALE）。ALE 是 ALD 的减材对应工艺，利用连续的自限性表面反应，以 Å 级精度去除材料。等离子体原子层刻蚀已在半导体加工中得到广泛应用，而热原子层刻蚀则通过机理研究和不断扩大的材料覆盖范围迅速发展 [34, 35, 23, 20]。在本节余下部分，我们将采用相同的基于 ORKG 的建模与查询方法，处理具有代表性的 ALE 综述表格。

为展示 ALE 的机器可操作建模，我们选取了六篇综述类文章，每篇都至少包含一个实质性表格，用于比较多项研究中的工艺条件、材料或刻蚀结果，并且设有专门的“Ref.”或“References”列。这类表格与 ORKG 出版模式十分契合：每一行都关联其来源文章，每一列则对应一个定义明确的属性。这些论文涵盖了等离子体和热原子层刻蚀，内容从一般性综述到针对特定材料的研究均有涉及。在以下小节中，我们将把这些表格转换为 ORKG 比较，并演示针对所报告刻蚀参数和结果的简单查询和复杂查询。

3.4 论文——原子层刻蚀：转折点上的综述

在这篇具有开创性的综述中，Oehrlein 等人（2015年） [43] 首次全面综述了原子层刻蚀，概述了关键概念、机制和材料系统，并比较了氧化物、III–V族半导体及相关基底中的卤素基和氟基化学体系。基于这项研究，我们将表I（“材料与原子层刻蚀研究综述”）建模为机器可操作的 ORKG 比较，以结构化形式收录其中调查的原子层刻蚀研究，作为构建原子层刻蚀知识图谱的起点。

3.4.1 表I的机器可操作建模

基于 Oehrlein 等人（2015年）表I中总结的原子层刻蚀调查问卷 [43], ，我们将其中报告的材料、化学体系和能源建模为 ORKG 比较 https://orkg. org/comparisons/R1562672 [21]。这一机器可读的比较针对每项原子层刻蚀研究，记录了刻蚀材料、吸附步骤中使用的前驱体化学以及驱动刻蚀或解吸的能源，从而为该表生成了结构化、可查询的表示。

底层数据涵盖氧化物、化合物半导体、氮化物、聚合物和碳基材料，展示了为实现原子尺度表面去除而开发的原子层刻蚀机制的多样性。

表18：关于 Oehrlein 等人表I的问题20的答案表。 [43], 在 ORKG 中建模为比较 R1562672。该表列出了每项原子层刻蚀研究所使用的材料、吸附前驱体化学和能源。

材料	前驱体化学	能源
Al_2O_3	BCl_3	氩中性束
BeO	BCl_3	氩中性束
GaAs	Cl_2	电子轰击
GaAs	Cl_2	248纳米氟化氪准分子激光器和钛蓝宝石激光器
InP	叔丁基膦	卤素灯解吸
Ge	Cl_2	电子回旋共振等离子体中的氩离子
石墨烯	O_2 等离子体	氩中性束
HfO_2	BCl_3	氩中性束
聚合物（聚苯乙烯）	O_2	来自CCP等离子体的氩离子
Si	Cl_2	来自ICP源的氩离子
TiO_2	BCl_3	氩中性束

简单查询。 第20个问题是：“对于表*I*中总结的每项原子层刻蚀研究，列出用于刻蚀或解吸的材料、吸附前驱体化学和能源。”此问题的 SPARQL 查询实现可通过 https://tinyurl.com/orkg-ale-materials-nolimit 重新运行。如前所述，本文中的答案是根据原始问题整理汇总的结果；表18列出了 ORKG 比较中记录的所有材料、前驱体化学和能源。

复杂查询。 第21个问题是：“按主要能源类别（中性束、等离子体离子、光子或热）对所有原子层刻蚀材料进行分组，并统计每个类别中使用的不同前驱体化学的数量。” 此问题的 SPARQL 查询实现可通过 https://tinyurl.com/ale-energy-classes 重新运行。与其他问题一样，答案是基于 ORKG比较进行的综合性机器生成聚合：所有原子层刻蚀条目均按能源类别分组，并返回每个类别中不同前驱体化学的数量。在当前比较中，等离子体或离子驱动的原子层刻蚀过程涉及12种不同的前驱体化学，中性束方法涉及四种，而光子辅助和其他活化方案各涉及一种。

3.5 论文——热原子层刻蚀：综述

Fischer等（2021年）的综述概述了热原子层刻蚀，重点介绍基于连续表面氟化和配体交换步骤的自限性无等离子体反应机制。该综述调查了氧化物、氮化物、半导体和金属领域的实验与理论研究，并讨论了其基本化学原理、反应器设计和机理分类。在其内容中，只有表III符合我们对 ORKG 中机器可操作建模的结构和溯源标准，因为该表按刻蚀材料和反应物化学系统列了原子层刻蚀出版物。因此，我们将表III转换为语义结构化的 ORKG 比较，以便对文献中的原子层刻蚀化学体系进行可复现、可查询的综合分析。

3.5.1 表III的机器可操作建模

综述文章 《热原子层刻蚀：综述》的表III按被刻蚀材料和反应物化学列出了原子层刻蚀文献。每一行对应一项实验报道的工艺，记录基底材料、主要和次要反应物以及文献参考信息。在机器可操作的形式中，该表在 ORKG 中被建模为比较 https://orkg.org/comparisons/R1563034 [4], ，其中每篇文献条目均表示为一个 compareContribution，包含四个关键谓词：P183144（被刻蚀材料）、P183145（反应物*1*）、P183146（反应物*2*）和 P183147（反应物*3*）。这种编码将文献摘要转化为可通过语义查询的数据集，为下文讨论的简单查询和复杂查询提供基础。

表19：Fischer等人表III中问题22的答案表， [20], 在 ORKG 中建模为比较R1563034。该表展示了每种材料类别的代表性材料、占主导地位的反应物元组（反应物1–3），以及使用这些元组的已报告 ALE工艺数量。

材料类别 / 代表性材料	占主导地位的反应物元组	# 报告数
氧化物（Al_2O_3, HfO_2, ZrO_2）	氢氟酸 + （三甲基铝、二甲基乙酰胺、Sn(acac)$_2$, $TiCl_4$)	25
氟化物 / 氮化物（AlF_3, AlN）	氢氟酸 + Sn(acac)$_2$	2
III–V族 半导体 (镓$_2O_3$, GaN, 铟镓砷)	氢氟酸/BCl_3 或氟化氙$_2$ 循环与二甲基乙酰胺/三甲基铝	4
过渡金属（钴、镍、铁、铜）	氧化/螯合：O_2/O_3 或Cl_2 + 六氟乙酰丙酮/乙酰丙酮	4
硅基体系（硅、SiO_2, Si_3N_4）	氢氟酸/三甲基铝、Cl_2，或 O_2/O_3 等离子体	6
难熔/导电薄膜（氮化钛、钨、WO_3, TiO_2）	组合 WF_6/BCl_3/O_2/O_3 配体交换路线	5
其他 / 新兴（VO_2, ZnO）	HF/SF_4 或 Hacac + O_2等离子体	3

简单查询。第 22 个问题是："对于每种被刻蚀材料，列出已报道的不同反应物元组（反应物*1–3*），以及报告每个元组的条目数。"此问题的 SPARQL 查询实现可通过 https://tinyurl.com/t3-easy-ale 重新运行。汇总整理后的答案见表19。该表列出了各材料类别中占主导地位的反应物元组，以及每种元组对应的已报道原子层刻蚀工艺数量。在此数据集中，该查询返回 46 个不同的材料—反应物元组。

表20：Fischer等人表III中问题 23的答案表 [20], ，在 ORKG 中建模为比较R1563034。对于每种机理原型，该表报告了不同原子层刻蚀化学体系的数量以及刻蚀材料的数量。

机制类型	# 化学体系	# 材料
氟化 + 配体交换	23	17
其他 / 未分类	11	10
氧化 + 螯合	3	4
卤化与转化	1	1

复杂查询。 第23个问题是："根据主导机理原型对每种 *ALE* 化学体系进行分类——*(i)* 氟化 +配体交换（*HF* 与 *TMA/DMAC/Sn(acac)$_2$*），*(ii)* 氧化 + 螯合（*O_2/O_3* 与 *Hacac/Hhfac*），*(iii)* 卤化与转化（*Cl_2/XeF_2* 与 *BCl_3/WF_6/WCl_6/$TiCl_4$*），或 *(iv)* 其他。" 此问题的 SPARQL 查询实现可通过 https://tinyurl.com/t3-mechanism-buckets 重新运行。如前所述，本文中的答案是基于原始问题整理并综合得出的结果；表20列出了 ORKG 比较中每种机理原型对应的不同化学体系和刻蚀材料数量。

3.6 论文——热原子层刻蚀：机理、材料与前景

Fang 等（2018年） [19] 发表了一篇广受引用的综述，介绍了热原子层刻蚀（ALE），并总结了其化学原理、机理路径和材料体系。文章对比了热原子层刻蚀与原子层沉积（ALD）和等离子体原子层刻蚀，并提出了机理分类，例如氟化–配体交换、转化–刻蚀以及氧化–氟化循环。文章还调查了材料与应用方面的挑战，包括温度依赖的 EPC 行为、前驱体设计和工艺选择性。在这一更广泛的背景下，表3对经实验验证的热原子层刻蚀工艺进行了跨材料总结，我们以此为基础构建机器可操作的 ORKG 比较。

3.6.1 表3的机器可操作建模

Fang等（2018年） [19] 中的表3汇总了多个材料类别中具有代表性的热原子层刻蚀工艺。对于每种工艺，该表列出了刻蚀材料、表面吸附反应物、表面

去除反应物、每循环刻蚀速率（EPC）和工艺温度。我们在 ORKG 中将此表建模为比较：https://orkg.org/comparisons/R1560222 [10], ，其中每种工艺均表示为一个独立的比较贡献，表格字段则编码为比较谓词。这种机器可操作表示为下文讨论的简单查询和复杂查询奠定了基础。

表21：Fang 等人表3中问题 24 的答案表， [19], 在 ORKG 中建模为比较R1560222。该表列出了所有 EPC 为 $>$ 05 埃/周期的热原子层刻蚀工艺，包括材料、表面吸附反应物、表面去除反应物、每周期刻蚀量和刻蚀温度。

材料	表面吸附	表面去除	每周期刻蚀量 [Å/周期]	温度 [°C]
W	O_2	WF_6	6.30	300
WO_3	BCl_3	HF	4.19	207
W	O_3/BCl 3	HF	2.50	207
ZnO	HF	TMA	2.19	295
AlN	HF	$Sn(acac)_2$/H_2 等离子体	1.96	250
HfO_2	HF	铝$(CH_3)_2Cl$	0.77	250
铝$_2O_3$	HF	$Sn(acac)_2$	0.61	250
TiO_2	WF_6	BCl_3	0.60	170

简单查询。 第24个问题是：“列出所有每周期刻蚀量大于 >0.5 埃/周期的热*ALE*工艺；返回材料、表面吸附反应物和表面去除反应物、每周期刻蚀量以及刻蚀温度；按每周期刻蚀量排序。”该问题的 SPARQL 查询实现可通过 https://tinyurl. com/t3-easy-fang 重新运行。ORKG 比较中每周期刻蚀量较高的工艺结果如表21所示。

表22：Fang 等人表3中问题 25的答案表。 [19], 在 ORKG 中建模为比较R1560222。对于每种机理原型，该表报告不同材料的数量以及平均每周期刻蚀量（单位：埃/周期）。

机理原型	# 材料	平均每周期刻蚀量 [Å/周期]
其他	2	5.245
氧化 + 氟化	2	1.350
氟化 + 配体交换	7	0.617
卤化 + 转化	1	0.600

复杂查询。 第25个问题是：“请按推断的机制类型对所有热原子层刻蚀工艺进行分组——*(i)* 氟化 + 配体交换（*HF* 与 *TMA/DMAC/Sn(acac)₂/Al(CH₃)₂Cl/SiCl₄*），*(ii)* 卤化 +转化（WF_6 与 BCl_3），*(iii)* 氧化 +氟化（O_2/O_3 与 *HF*），或 *(iv)* 其他——并计算每组中不同材料的数量和平均每循环刻蚀量。”该问题的 SPARQL 查询实现可通过 `https://tinyurl.com/t3-complex-fang`重新运行。汇总结果如表22所示。

3.7 论文 – 定向原子层刻蚀的物理和化学效应

Sang和Chang（2020年） [56] 综述了决定定向原子层刻蚀的物理和化学因素，重点阐述离子–表面相互作用、能量阈值、溅射、表面氧化、配体交换和反应热力学如何相互作用，从而产生各向异性和选择性。文章讨论了基于等离子体、热驱动以及等离子体–热混合的原子层刻蚀途径，并将离子驱动现象和表面反应的基本原理与先进半导体图形化的集成要求联系起来。

3.7.1 表1的机器可操作建模

Sang和Chang（2020年）中的表 1 总结了适用于元素半导体和化合物半导体的代表性 ALE 工艺，包括硅、锗、砷化镓、磷化铟、氮化镓、铟镓砷和氮化铝镓。对于每种工艺，该表记录了刻蚀材料、活化机制（通常由等离子体驱动）、表面改性化学、去除步骤以及关键工艺条件。该表还包括两种纯

热原子层刻蚀系统：氮化镓（氟化氙/BCl_3）和铟镓砷（氢氟酸–吡啶/$AlCl(CH_3)_2$）。我们将此表建模为 ORKG 中的比较：https://orkg.org/comparisons/R1560825 [63]。这种机器可操作表示是下文将讨论的简单查询和复杂查询的基础。

表 23：关于 Sang和Chang 的表 1 的问题26答案表 [56], , 在 ORKG 中建模为比较 R1560825。该表列出了半导体原子层刻蚀工艺及其改性、去除和活化模式。

半导体	改性	去除	活化
氮化铝镓	Cl_2	Ar	等离子体
GaAs	Cl_2	Ar	等离子体
GaN	Cl_2	Ar	等离子体
Ge	Cl_2	Ar	等离子体
铟镓砷	Cl_2	Ar	等离子体
InP	Cl_2	Ne	等离子体
Si	Cl_2	Ar	等离子体
GaN	XeF_2	BCl_3	热
铟镓砷	HF-吡啶	铝$Cl(CH_3)_2$	热

简单查询。 第 26 个问题是：“列出所有半导体原子层刻蚀工艺，并分别返回其改性、去除和活化类型；按活化模式排序。”此问题的 SPARQL 查询实现可通过 https://tinyurl.com/t1-semi-ale 重新运行。所得半导体原子层刻蚀工艺见表 23。

表 24：针对 Sang和Chang 的表 1 [56], （在 ORKG 中建模为比较 R1560825）的 Q.27 答案表。该表按活化类型列出改性–去除组合，以及采用各组合的不同半导体材料数量。

活化	改性 / 去除组合	# 半导体
等离子体	Cl_2 / 氩气	6
等离子体	Cl_2 / 氖	1
热	HF-吡啶 / 铝$Cl(CH_3)_2$	1
热	氟化氙 / BCl_3	1

复杂查询。 第27个问题是：“按活化类型（等离子体与热）对半导体原子层刻蚀工艺分组，并统计采用每种改性–去除对的不同材料数量。”此问题的 SPARQL 查询实现可通过以下链接重新运行： `https://tinyurl.com/t1-sang-complex`。汇总结果见表 24。

3.8 论文——金属的各向异性/各向同性原子层刻蚀

Kim 等人 (2020年) [55] 回顾了金属原子层刻蚀的最新进展，对比了各向异性的离子辅助路径与纯热式各向同性路径。鉴于先进半导体器件和环绕栅极、垂直场效应晶体管等三维架构面临的尺寸缩小挑战，该文将原子层刻蚀定位为传统反应离子刻蚀的一种低损伤、埃级替代方案。文章调查了已报道的金属和金属氮化物（如铜、钨、钴、铬、铁、铂、氮化钛、氮化铝）的原子层刻蚀化学体系，并讨论了吸附步骤、离子能量窗口、热反应和工艺理想性。在此背景下，表2总结了金属原子层刻蚀系统的方向性、反应物组合、工艺温度和每周期刻蚀量，并为我们的机器可操作的 ORKG 比较提供依据。

3.8.1 表 2 的机器可操作建模

Kim等人（2020年） [55]总结了已报道的金属和金属氮化物 ALE 工艺。对于每条记录，表中列出了刻蚀材料、方向性（各向异性或各向同性）、表面改性和去除反应物、工艺温度以及每周期刻蚀量。涉及的材料包括 Co、Fe、Pd、Pt、Cr、W、Cu，以及 TiN 和 AlN 等氮化物。我们将此表建模为 ORKG 中的比较 https://orkg.org/comparisons/R1563131 [12], , 形成一个机器可操作的数据集，支持下文介绍的简单查询和复杂查询。

表25：关于 Kim 等人的表2中问题28的答案表， [55], 在 ORKG 中建模为比较R1563131。该表列出了金属原子层刻蚀工艺及其每 $\geq$ 2 /周期的每周期刻蚀量，包括方向、改性和去除化学反应、工艺温度以及循环时间。

材料	方向	反应去除		EPC	温度（°C）	循环时间（秒）
Fe	各向异性	O_2	甲酸	42	80	350
Co	各向异性	O_2	甲酸	28	80	350
Pd	各向异性	O_2	甲酸	12	80	350
W	各向同性	O_2	WF_6	6.3	300	320
Pt	各向异性	O_2	甲酸	5	80	350
W	各向异性	NF_3	O_2	2.6		40
W	各向同性	O_2/O_3	BCl_3→氢氟酸	2.5	207	325
W	各向异性	Cl_2	Ar	2.1	60	25
Co	各向同性	Cl_2	六氟乙酰丙酮	2	140	95

简单查询。第28个问题是：“列出所有每周期刻蚀量为 ≥ 2 / *cycle* 的金属原子层刻蚀工艺；返回材料、方向（各向异性或各向同性）、改性和去除化学体系、每周期刻蚀量、工艺温度和循环时间；按每周期刻蚀量降序排列。”此问题的 SPARQL 查询实现可通过以下链接重新运行：
`https://tinyurl.com/t2-metals-high-epc`。结果显示的高每周期刻蚀量金属原子层刻蚀工艺见表25。

表26：基于 Kim 等人的表2回答问题 29，该表在 ORKG 中建模为比较R1563131。对于每个刻蚀方向，表中报告了不同金属的数量和平均每循环刻蚀量（单位：埃/周期）。

方向	金属数量	平均每循环刻蚀量（埃/周期）
各向异性	7	11.6875
各向同性	5	2.035

复杂查询。 第29个问题是：“按方向（各向异性与各向同性）对金属原子层刻蚀工艺进行分组，并计算每组涵盖的不同金属数量及平均每循环刻蚀量；按平均每循环刻蚀量降序排列。”此问题的 SPARQL 查询实现可通过以下链接重新运行： `https://tinyurl.com/t2-metals-complex`。汇总结果见表26。

3.9 论文——纳米尺度半导体器件的 SiO_2 原子层刻蚀：综述

Hong等人（2024年） [27] 综述了先进半导体图形化中的 SiO_2 原子层刻蚀策略。他们比较了含氟碳化合物的各向异性原子层刻蚀——其受前驱体选择、离子能量、刻蚀选择性和腔壁效应控制——以及各向同性原子层刻蚀路线；后者在热激活或等离子体激活下，将 SiO_2 转化为 Al_2O_3 或氟硅酸铵等挥发性中间产物。通过将机理步骤与器件相关工艺指标相联系，该综述探讨了如何调整 SiO_2 原子层刻蚀，以满足新兴三维架构和亚10纳米制造要求。

3.9.1 表 I–VI 的机器可操作建模

SiO_2 原子层刻蚀综述中的表 I–VI 涵盖了含氟碳化合物的各向异性原子层刻蚀，以及热激活或等离子体辅助的各向同性机制。表 I 和表 II（https://orkg.org/comparisons/ R1560949 [28], https://orkg.org/comparisons/R1560977 [29]）列出了各向异性 SiO_2 原子层刻蚀所用的氟化前驱体、去除气体、刻蚀速率、温度和离子能量窗口。表 III 和表 IV（https://orkg.org/comparisons/R1561025 [30], https://orkg.org/ comparisons/R1561023 [31]）总结了选择性（SiO_2/Si 和 SiO_2/Si_3N_4）、提高选择性的方法以及腔壁调理效应。表 V 和表 VI（https://orkg.org/ comparisons/R1561046 [32], https://orkg.org/comparisons/R1562478 [33]）介绍了热激活和等离子体辅助的各向同性原子层刻蚀，并将多步反应途径与刻蚀速率一并归纳。

与本研究建模的其他原子层刻蚀表格一样，全部六个 SiO_2 表格均在 ORKG 中表示为采用共享统一谓词模式的机器可处理比较。对于表 I–

IV，核心描述符包括 P183129（氟化前驱体化学）、P183125（去除化学）、P183130（工艺温度）和 P183131（刻蚀速率）；表 II 还增加了与等离子体相关的 P183132（去除步骤中的离子能量）。表 III 引入了侧重选择性的谓词 P183133（选择性材料对）、P173032（选择性数值）和 P183134（选择性改进方法），而表 IV 则提供 P183135（腔壁效应的方法）。各向同性原子层刻蚀机制表（V 和 VI）使用 P183136（刻蚀机制类别）以及步骤级谓词 P183137、P183138 和 P183139。在这一共享模式下，跨表 SPARQL 查询能够在整个原子层刻蚀语料库中关联前驱体系列、离子能量窗口、选择性范围、腔壁处理和各向同性机制路径。

这六项 SiO_2 ALE 比较也收录在 ORKG 智能综述（https://orkg. org/reviews/R1562498）中，该综述为 Hong等人（2024年） [27]建模的表格提供了统一入口。虽然可以直接对这些比较执行跨表 SPARQL 查询，但该智能综述还记录了建模选择、展示了示例查询，并将语义数据模型与原始文章关联起来。

表 27：问题3.30（简单跨表查询 1）的答案表，基于 Hong 等人 [27],的表 I 和表 II，这两张表在 ORKG 中分别建模为比较 R1560949 和 R1560977。该表列出了两张表中均出现的氟碳原子层刻蚀工艺，这些工艺采用接近室温（RT ±20°C）条件，并列出其去除化学、刻蚀速率和离子能量窗口。

前驱体	去除	温度（°C）	每周期刻蚀量（Å/周期)	离子能量窗口
C4F6/Ar 等离子体	氩等离子体	−10	14.2	10–100 V
氟化亚甲基$_3$ 等离子体	氩等离子体	−40–20	9.0	0–50 V
C4F8等离子体	氩等离子体	20	5.8	30–90 V
$CF_3CF_2CF_2CH_2OH$ 等离子体	氩等离子体	RT	5.2	10–80 V
氟化亚甲基$_3$ 等离子体	O_2 等离子体	20	4.1	30–90 V
n-$C_3F_7OCH_3$ 等离子体	氩等离子体	RT	2.1	10–80 V
C4F8等离子体	氩离子束	RT	1.9	30–200 伏
i-$C_3F_7OCH_3$ 等离子体	氩等离子体	RT	1.8	10–80 伏

简单跨表查询 1. 第30个问题是：“对于同时出现在表*I*和表 *II*中的每种氟碳前驱体系统，列出去除步骤所用的前驱体化学、去除气体、工艺温度、刻蚀速率和离子能量窗口；将条件限制在接近室温（*RT* ±20°*C*）的范围内，并按刻蚀速率降序排列。”此问题的 SPARQL 查询实现可通过以下链接重新运行：`https://tinyurl.com/pap6-crossq1-easy`所得氟碳原子层刻蚀工艺见表 27。

表 28：问题 31（简单跨查询2）的答案表，基于 Hong 等人的表 III 和表 IV， [27],在 ORKG 中建模为比较 R1561025 和 R1561023。该表列出了两张表中均出现的基于 C_4F_8/Ar 的各向异性硅氧化物原子层刻蚀工艺，包括其选择性配对、选择性范围、选择性测定方法、腔室壁处理和刻蚀速率。

选择性 pair	选择性	选择性测定方法	腔室壁处理	每周期刻蚀量（埃/周期）
硅O_2/硅$_3N_4$	0.2–15.0	前驱体选择；含氧碳膜厚度；离子能量；刻蚀步骤 time	腔室清洁 with O_2 等离子体；腔室 wall 供暖	3
硅O_2/硅$_3N_4$	$>$ 7.0	前驱体选择	腔室清洁 with O_2 等离子体；腔室 wall 供暖	3
SiO_2/Si	$>$ 10.0	前驱体选择	腔室清洁 with O_2 等离子体；腔室 wall 供暖	3

简单跨查询2. 第31个问题是：“汇总表*III*和表 *IV*中所有基于 C_4F_8/氩等离子体的各向异性 SiO_2 *ALE* 工艺，并分别列出目标选择比对（SiO_2/硅或 SiO_2/硅$_3N_4$）、选择性范围（如有）、腔室壁处理（无、O_2 清洁、腔壁加热）和刻蚀速率；按刻蚀速率排序。”此问题的 SPARQL 查询实现可通过以下链接重新运行：`https://tinyurl.com/pap6-crossq2-easy`匹配的工艺见表 28。

表 29：针对 Hong 等人的表 I–III，对问题3（复杂跨查询1）的回答表 [27],，建模于 ORKGas 比较 R1560949、R1560977 和 R1561025 中。对于每种含氟碳前驱体族，该表列出了 SiO_2 刻蚀速率的平均值和最大值、合并离子能量窗口，以及报告的 SiO_2/Si 和 SiO_2/Si_3N_4 选择性的最大值。

前驱体类别	平均刻蚀速率 rate	最大刻蚀速率 rate	离子能量窗-dows	最大选择比 SiO_2/硅	最大选择比 SiO_2/硅$_3N_4$
C_4F_8	—	—	—	0	0
$C_3H_3F_3$	—	—	—	0	0
C_4F_6	—	—	10–100 V	—	—
CF_3I	—	—	—	—	—
CHF_3	—	—	—	0	0

复杂跨查询1。 第32个问题是：“跨表*I*、表*II*和表*III*，按含氟碳前驱体族对所有各向异性 SiO_2 *ALE*工艺进行分组（例如，基于C_4F_8、基于CHF_3、$C_3F_7OCH_3$ 异构体）。对于每个类别，计算：*(i)* 平均和最大SiO_2 刻蚀速率；*(ii)* 去除步骤中所报告的*ALE*离子能量范围的并集；以及 *(iii)* 所报告的最大SiO_2/*Si*和SiO_2/Si_3N_4 选择比；每个前驱体类别返回一行，先按最大选择比排序，再按平均刻蚀速率排序。” 该问题的SPARQL查询实现可通过以下链接重新运行：`https://tinyurl.com/pap6-crossq1-complex`。结果得到的前驱体类别聚合结果见表 29，反映了当前建模表格中可解析的数值刻蚀速率和选择比数据仍然有限。

表 30：Hong 等人表 V 和 VI 的问题 33（复杂交叉查询 2）答案表， [27],在 ORKG 中建模为比较 R1561046 和 R1562478。各向同性 SiO_2 原子层刻蚀工艺按刻蚀机制类别分组；对于每个类别，表中列出了变体数量、刻蚀速率的最小值/最大值/平均值、与限速步骤相关的温度范围，以及是否需要等离子体、是否实现低温高速运行。

机制类别	变体数量	最低速率	最高速率	平均速率	温度范围 [°C]	需要等离子体	高速率、低-T
基于 AFS 的等离子体辅助	3	27	75	43	3–4	yes	yes
基于 AFS 的热过程	1	9	9	9	20–20	no	yes
Al_2O_3-转化循环	2	0.07	0.35	0.21	300–350	no	no

复杂交叉查询 2（表 V +VI：热辅助与等离子体辅助各向同性原子层刻蚀）。第 33 个问题是：“汇总所有各向同性硅氧化物原子层刻蚀工艺，并按机制类别分组；报告变体数量、刻蚀速率的最小值/最大值/平均值、温度范围以及是否需要等离子体。”此问题的 SPARQL 查询实现可通过以下链接重新运行：`https://tinyurl.com/pap6-crossq2-complex-fixed`汇总后的机制类别统计数据见表 30。

至此，我们结束关于 ORKG 中 ALE 机器可处理比较的章节。

在本研究考察的所有 ALD 和 ALE 综述论文中，我们将共计**18 个机器可处理的表格**建模为 ORKG 比较：其中**七个**来自三篇 ALD 论文，另有**十一个**来自六篇 ALE 论文。这些结构化表示为前驱体和反应物化学体系、工艺参数、刻蚀与生长指标、选择性行为、离子能量窗口、腔壁效应以及各向同性多步机制提供了统一的语义层。总计**33 个自然语言问题**——其中 ALD 有**19 个**，ALE 有**14 个**——均已实现为 SPARQL 查询，每个查询都会生成可复现、由机器得出的结果表。这些查询共同构成了一个全面的基准测试，用于在 ORKG 中对材料科学知识进行神经符号查询。

作为本文的补充数据，我们在 GitHub 仓库中提供了一份汇总目录，其中包含论文、表格、比较、自然语言问题和 SPARQL 链接：https://github.com/ sciknoworg/ald-ale-orkg-review。

在本文的其余部分，我们以 SPARQL 查询结果作为黄金标准答案，在该数据集上评估了几种基于大语言模型的问答设置。

4 评估与结果

在为材料科学综述文章中选定部分发布所有机器可操作的符号表示后，我们接下来进行定量评估，以回答文章开篇提出的三个研究问题：（i）领域专家认为这些基于 ORKG 的比较表示有多大用处（**RQ1**）；（ii）纯神经方法、基于 PDF 的大语言模型查询与符号化 SPARQL 基线相比表现如何（**研究问题2**）；以及（iii）相较于前两者，针对 ORKG 比较进行神经符号查询的表现如何（**RQ3**）。这些评估共同证明了符号化科学知识表示对于促进神经符号人工智能的重要性——这也是本文的核心研究主张。

```sparql
PREFIX orkgr: <http://orkg.org/orkg/resource/>
PREFIX orkgp: <http://orkg.org/orkg/predicate/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT
  (COALESCE(?reactorLabel, STR(?reactor)) AS ?reactor_value)
  (COALESCE(?lharLabel, STR(?lhar))       AS ?lhar_value)
  (COUNT(*) AS ?n)
WHERE {
  orkgr:R1469158 orkgp:compareContribution ?contrib .
  ?contrib orkgp:P180007 ?reactor .
  ?contrib orkgp:P180003 ?lhar .
  OPTIONAL { ?reactor rdfs:label ?reactorLabel . }
  OPTIONAL { ?lhar    rdfs:label ?lharLabel . }
}
GROUP BY ?reactor ?reactorLabel ?lhar ?lharLabel
ORDER BY DESC(?n) ?reactor_value ?lhar_value
```

⟺

```
Please extract all combinations of reactor
types and LHAR structures from Table 2 of
the article titled "Saturation profile based
conformality analysis for atomic layer
deposition: aluminum oxide in lateral
high-aspect-ratio channels". Count how many
times each combination occurs across the
included studies. The result must be
returned as one single table only, in CSV
format, with one row per result. The table
should have the following columns:
reactor_value, lhar_value, n.
```

图 6：SPARQL 查询作用于机器可操作的 ORKG 比较图 https://orkg.org/comparisons/R1469158 及其自然语言等价形式的示意图。后者可用于针对不同形式输入（如文章 PDF 或比较的序列化版本）进行基于大语言模型的查询。此 SPARQL 查询在第 3.1 小节中介绍。

4.1 评估设置

我们首先介绍评估设置：第 4.1.1 小节介绍实验数据集，第 4.1.2 小节概述实验设置。

4.1.1 实验数据集

我们的评估基于第 3 节介绍的 33 个 SPARQL 查询：其中 19 个问题针对 ALD 比较，14 个问题针对 ALE 比较。每个查询在一个或多个 ORKG 比较上运行，并返回一个机器生成的答案表，我们将其用作金标准。

对于神经网络（基于大语言模型）的实验，我们随后为每个 SPARQL 查询创建了一个自然语言版本。这些自然语言查询明确指定了：（i）应查阅论文中的哪个综述表格；（ii）应使用该表格中的哪些列；以及（iii）答案表应包含哪些列，并要求将输出格式化为表格。换言之，它们传达的信息与相应的 SPARQL 查询相同，但形式上能够让大语言模型直接遵循。图 6 展示了一个示例。我们使用 mistral-large 以编程方式生成了这些自然语言查询；Python 脚本位于我们的 GitHub 仓库中：https://github.com/sciknoworg/ald-ale-orkg-review/blob/main/llm-experiments/nl_query_generator_csv_onetable.py。由此生成了 33 个详细的自然语言查询，它们分别与 33 个 SPARQL 查询及其金标准答案表一一对应。请注意，这些查询不同于前文（第 3 节）较为非正式的问题表述；后者面向读者，而不是用于强制大语言模型生成结构一致的输出。为便于快速查阅，GitHub 仓库还针对每个 ALD 和 ALE 综述表格（包括跨表查询）列出了简短版和详细版自然语言查询，以及对应的 SPARQL 查询链接。

对于其中一种实验设置（下文所述的设置 3），大语言模型看不到完整的综述 PDF。（相比之下，在设置 2 中，模型会获得综述 PDF 或其提取出的文本，并且

必须阅读这些内容才能回答问题。）在设置 3 中，为了使神经网络实验在理念上更接近设置 1 的纯 SPARQL（符号）金标准，我们改为以简单的类电子表格格式（CSV 文件）提供 ORKG 比较表本身。该 CSV 中的每一行对应原综述表格中的一行，每一列对应我们在 ORKG 中建模的一项类型化属性（例如材料、前驱体化学、基底、温度、刻蚀速率或选择性）。换言之，模型看到的结构化信息与设置 1 中 SPARQL 查询所处理的信息相同，只不过现在它呈现为平面文本表格，而非嵌入 PDF 中的图。我们通过 ORKG 内置的导出功能获取这些 CSV 文件；该功能还支持其他格式，例如 LaTeX、PDF、RDF 和引文元数据。因此，设置 2 和设置 3 都是通过大语言模型评估问答能力，与设置 1 中直接进行结构化图查询形成对比。SPARQL 查询具有完全确定性，对于给定的 ORKG快照，总会返回相同结果；而大语言模型是概率模型：即使底层证据（此处为精选比较表）保持不变，其回答在不同运行中也可能略有差异。因此，设置 2 和设置 3 中的回答与 SPARQL 答案高度一致，应被视为对符号金标准的近似，而非精确保证。

4.1.2 实验设置

设置 *1*：对结构化比较表进行符号查询。在这一设置中，我们完全不使用任何语言模型，而是直接使用 SPARQL（一种用于图的结构化查询语言）查询机器可操作的 ORKG 比较表。对于 33 个问题中的每一个，我们都编写了相应的 SPARQL 查询，从精选比较表中准确选取相关的行和列。由于这些查询针对结构化数据运行且执行过程具有确定性，因此生成的答案表构成了金标准参照，用于与所有基于大语言模型的设置（设置2和设置3）进行比较。

设置 *2*：从基于 *PDF* 的科学证据中进行神经网络查询。在这一设置中，我们考察大语言模型仅以综述文章的 PDF 版本作为证据时，回答这 33 个问题的表现如何。专有模型和开放权重模型都会收到 SPARQL 查询的详细自然语言等价形式，以及文章 PDF（或从中提取的文本和表格），但它们无法直接访问 ORKG 表示。因此，模型必须通过单次神经推理，内部解析 PDF、定位相关段落和表格，并生成答案表。将模型输出与设置 1 中的 SPARQL 金标准进行比较后，我们便能量化纯神经网络的 PDF 查询在回答精确的表格型科学问题时，能在多大程度上接近符号方法的性能。

2.1 专有语言模型（神经网络）查询。我们首先评估三种最先进的商业托管对话模型——ChatGPT-5.1（OpenAI，发布于 2025年11月12日） [47],Gemini-3-Pro（Google，发布于 2025 年 11 月 18 日） [50], 和 Claude-Sonnet-4.5 （ Anthropic，发布于 2025 年 9 月 29日） [5]。截至本文提交日期（2025年12月13日），这些系统属于能力最强的专有大语言模型之列。我们向每个模型提供 SPARQL 查询的详细自然语言等价形式，以及相应综述文章的 PDF。此设置旨在检验这类黑箱系统能否仅凭 PDF，借助其自身的 PDF 解析流程，直接回答精确且面向表格的科学问题。此设置的局限在于，无法检查或修改解析和上下文整合步骤，且除服务提供商在其界面中公开的信息外，无法精确复现运行过程。

2.2 开放权重语言模型（神经）查询。 为实现充分的实验控制和可复现性，我们在专有系统之外，还采用了性能出色的开放权重大语言模型。与通过付费云服务访问的专有模型不同，这些开放权重模型公开其参数，可从 https://huggingface.co/models 或 Ollama https://ollama.com/ 等平台下载并在本地运行。只要具备合适的硬件，就无需按查询付费，同时还能全面掌控模型版本管理和配置，严格保证实验的可复现性，并可选择将所有数据保留在本地计算环境中（这对于敏感数据集可能尤为重要）。

对于每篇论文，我们首先使用 pdfplumber 和我们的处理脚本 https://github.com/sciknoworg/ald-ale-orkg-review/blob/main/llm-experiments/pdf_extractor.py 将 PDF 转换为机器可读的文本和表格。具体而言，pdfplumber 读取每个 PDF 页面，并提取可见文本以及表格的网格

结构，将其转换为纯文本和表格形式。这一步转换是必需的，以便语言模型能够将文章内容作为文本标记序列以及行/列进行处理，而不是处理一种针对显示进行了优化、但其信息无法直接用于文本处理的页面布局格式。随后，我们评估五种开放权重模型：gemma-3-27b-it、llama-3.3-70b-instruct、qwen2.5-vl-72b-instruct、qwen3-30b-a3b-instruct-2507 和 mistral-large-instruct；这些模型是根据其在近期表格推理基准中的出色性能选定的 [11]。

对于开放权重模型，我们考虑两种变体：

a) 全文档上下文（上下文注入式文档问答）。在这种变体中，给定论文中提取的所有文本和表格会被拼接成一个长上下文，并与详细的自然语言查询一同直接传递给模型（参见我们存储库中的 context_injected_doc_qa.py）。模型必须根据这一完整但可能含有噪声的文章表示作答。这模拟了人类通读整篇论文后回答问题的方式。

b) 基于 *PDF* 片段的检索增强生成（*RAG*）。在检索增强生成变体中，我们首先将提取的论文拆分为连续的文本段（每段大约包含几页内容），并使用稠密语义嵌入模型为这些文本段建立索引。在查询时，我们会检索与详细自然语言查询最相关的文本段，并仅将这些检索片段及查询提供给模型（参见 rag_pdf_segments_doc_qa.py）。这样可以缩短上下文长度，使模型专注于最可能包含答案的论文部分，但如果关键表格行或说明分散在相距较远的文本段中，就可能遗漏信息。

附录 A 中提供了设置 2 描述的详细技术版本，其中包括我们在 PDF 提取、分段、嵌入和检索超参数方面的实现选择。

设置 *3*：符号上下文增强的神经查询。在此设置中，我们通过将语言模型直接置于机器可操作的 ORKG 比较表中，而非原始 PDF 中，使其尽可能接近符号基线。具体而言，我们将每个 ORKG 比较导出为扁平的、类似电子表格的格式（CSV），如第 4.1.1 节所述，并将该 CSV 与详细的自然语言查询一同作为模型的输入。这些数据已公开发布 [15]。我们评估与设置 2 中相同的专有模型和开放权重模型，但在此设置中，它们完全看不到 PDF 或从中提取的文本。所有推理均基于 CSV 编码的比较表进行，这些比较表在内容上与设置 *1* 中通过 *SPARQL* 查询的比较表完全相同；唯一变化的是查询引擎，从确定性的符号执行器变为概率性的神经模型。

我们将这种方式称为符号上下文增强查询，因为语言模型获得的不仅是一个自然语言问题，还包括证据的符号表示：ORKG 比较表。在人工智能和数据科学中，“符号”通常指以显式符号和关系表示的知识（例如表格中的行和列，或图中的节点和边），这些知识可通过 SPARQL 等形式语言进行查询，而非作为自由生成的文本。在我们的案例中，每种材料、前驱体、温度、刻蚀速率或外量子效率值都出现在具有明确定义含义的专用列中，而不是嵌入 PDF 的句子里。因此，设置 3 使我们能够研究：当语言模型的输入上下文包含这种结构化、类似符号的表格表示，而非仅包含异构的 PDF 布局时，它们能在多大程度上逼近 SPARQL 金标准。不过，我们并未构建专门的神经符号架构（例如在模型内部设置独立的符号推理模块）；符号信息纯粹是通过更丰富的输入上下文提供的。

需要注意的是，设置 1 中的查询引擎完全确定：对于固定的 ORKG快照，同一个 SPARQL 查询始终会返回相同的结果。相比之下，设置2和设置3中的神经查询引擎具有非确定性，因为其文本生成涉及随机选择；即使输入不变，重复运行也可能得到略有不同的答案。在设置 3 中，模型仍必须执行 SPARQL 明确完成的关键操作：识别对比表中相关的行（例如，所有具有指定材料或前驱体的条目），选择符合所需输出的正确列，有时还要比较或合并不同行中的数值。因此，我们的评估旨在考察符号化落地的神经问答系统在多大程度上能够达到 SPARQL 基线（设置 1）的数值准确性，同时提供灵活的自然语言前端。

针对所有 ALD 和 ALE 问题，我们将 33 个查询应用于 21 种不同的模型—设置配置（例如，在设置 2a、2b 和 3 中分别评估每个模型），共进行了 693 次独立实验。这为比较符号查询、神经查询和符号上下文增强查询提供了相当充分的依据。

4.1.3 评估指标

我们采用了与三个研究问题相对应的不同评估指标。

专家对 SPARQL 查询的判断（RQ1）。为回答 RQ1，我们邀请了三位原子层沉积/原子层刻蚀领域专家开展了一项小规模调查（其中两位是本文的共同作者）。针对 33 个查询（19 个原子层沉积查询、14 个原子层刻蚀查询），专家使用提供的 ORKG 链接和 SPARQL 链接，在相应的 ORKG 比较中运行 SPARQL 查询，检查生成的表格，然后从两个维度对其进行评分：

- 有意义性：这个问题在科学上是否合理（例如，是否询问合理关系、使用恰当的概念，并符合当前对原子层沉积/原子层刻蚀的认识）？
- 实用性：生成的表格是否能为研究人员提供有帮助的综合结果（例如，节省时间、呈现难以通过目视发现的模式，或提供超出通读一次综述所能获得的非平凡见解）？

两个维度均采用 5 点李克特量表评分（1 = 非常不同意， 5 = 非常同意）。专家评估了涵盖七个原子层沉积比较表和十一个原子层刻蚀比较表的全部 33 个查询。对于每个查询，他们先判断自然语言问题本身是否有意义，然后判断相应的 SPARQL 结果表是否可作为综合工具发挥作用，前提是未来可以持续纳入更多研究并更新该表。将有意义性与实用性分开评估，有助于判断是否需要改进：(i) 问题的表述方式；(ii) 比较中底层数据的建模方式；或 (iii) 综合结果的呈现方式。对于实用性评分低于 3 的情况，我们请专家在自由文本字段中简要说明评分理由，以便收集有关局限性或缺失信息的定性反馈。调查问卷可在 https://forms.gle/qSDUG1WzSZSz2864A 获取；为确保透明度，我们将其公开，欢迎感兴趣的读者查阅或提交自己的回答。

神经网络表格与 SPARQL 表格的一致性（研究问题2 和RQ3）。为了解答研究问题2 和RQ3，我们逐查询衡量神经系统生成的表格与 SPARQL 结果表的匹配程度。为此，我们采用相对映射相似度 (RMS)指标。该指标由 Liu 等人 [38], 提出，最初用于评估表格与图表问答系统。

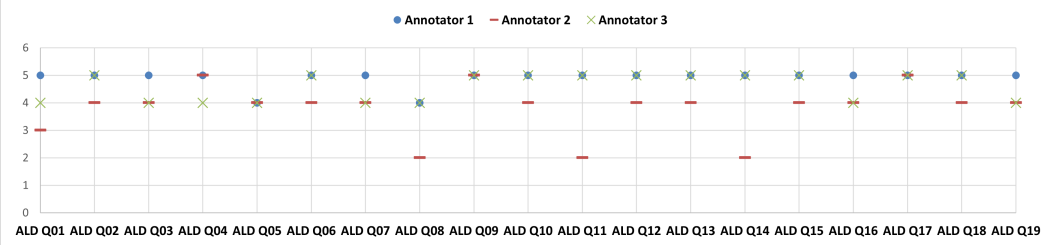
直观而言，RMS 将每个表格视为由 (行标题、列标题、单元格值) 这类条目组成的集合，并通过匹配这些条目来比较预测表格和黄金表格。只有在 (i) 模型选中了正确的行和列（例如，正确的材料或工艺），且 (ii) 对应的数值（例如，温度、刻蚀速率或外量子效率）接近 SPARQL 表格中的数值时，才能获得较高的 RMS 分数。RMS 会生成取值范围为 $[0, 1]$, 的精确率、召回率和 F1 分数，并且其设计使其不受行列顺序或表格转置的影响。

在我们的实验中，对于每个查询，我们将 SPARQL 结果视为黄金表格，并将每个神经系统（基于 PDF 和符号上下文增强设置）的输出视为预测。我们使用 Google Research 提供的官方 DePlot 实现（针对我们的 ALD/ALE 数据集进行了适配）计算 RMS 精确率、召回率和 F1 值；RMS 评估脚本和配置记录于 https://github.com/sciknoworg/ald-ale-orkg-review/blob/main/ eval_deplot_rms/README.md。RMS 的完整数学定义及实现细节见附录 B。

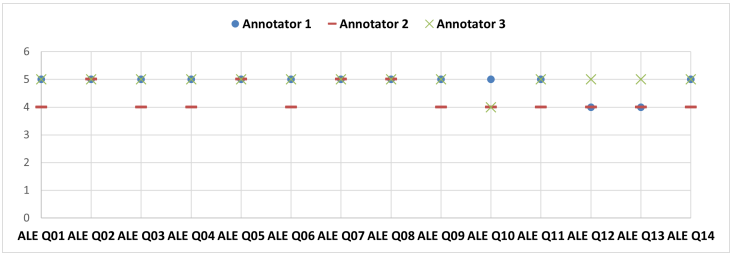
4.2 结果

4.2.1 领域专家对机器可操作知识综合的印象

在此，我们回答 RQ1：ALD/ALE 专家认为基于 ORKG 的比较表在回答复杂综述问题方面有多大用处。如评估指标章节 4.1.3 所述，三位

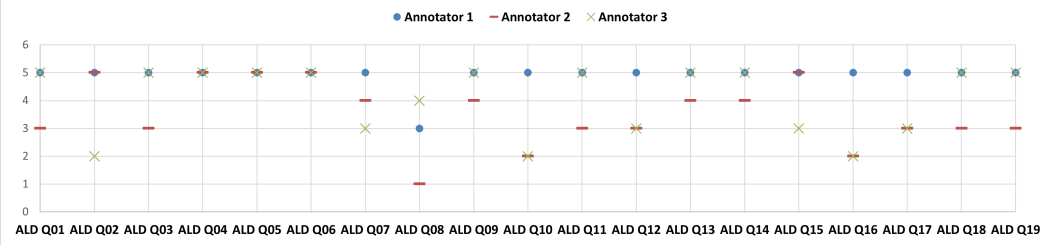


(a) 19 个 ALD 查询的逐项有意义性评分。

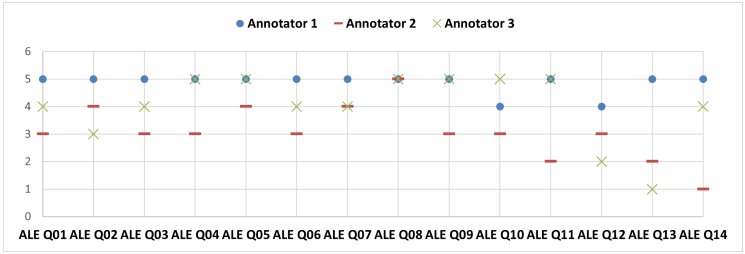


(b) 14 个 ALE 查询的逐项有意义性评分。

图 7：三位领域专家对所有 ALD（a）和 ALE（b）查询的有意义性评分，采用 5 点李克特量表（1 = 非常不同意， 5 = 非常同意），针对提示语“该查询有意义”作答。



(a) 19 个 ALD 查询的逐项有用性评分。



(b) 14 个 ALE 查询的逐项有用性评分。

图8：三位领域专家对所有 ALD（a）和 ALE（b）查询的有用性评分，采用 5 点李克特量表（1 = 非常不同意， 5 = 非常同意），针对提示语“查询结果提供了对表中见解的有用综合”作答（并要求在原始数据稀疏时，设想该表是内容更丰富且持续更新的版本）。

ALD/ALE 领域专家（两位共同作者和一位外部资深研究人员，他们在材料科学领域拥有 1–5 年及> 10 年的经验）完成了一份在线调查问卷，并在其中评估了全部 33 条查询（19 条 ALD，14 条 ALE）。对于每条查询，他们都在相应的 ORKG 比较上运行了 SPARQL 查询，检查生成的表格，并评定 (i) 该问题是否具有科学意义，以及 (ii) 该表格是否提供了有用的综合。

总体而言，反馈明显呈积极态势。专家认为几乎所有查询在科学上都具有良好定义：所有查询的平均有意义性评分为 $4.54/5$，其中 96% 的评分

为 ≥ 4，且 33 个查询中有 32 个的平均得分至少达到 4 分。图 7 展示了 ALD 和 ALE 各查询有意义性评分的分布。实用性评分略低，但仍然较为理想（平均值为 $4.08/5$）；33 个查询中有 25 个的平均实用性至少达到 4 分，只有两个查询低于 3 分；图8展示了相应的实用性评分。在关于机器可操作的科学知识价值的另一项总体问题中，三位专家的评分均介于 4 分和 5 分之间（平均值为 4.67）。换言之，专家不仅认为这些问题具有意义，也认为由 SPARQL 得出的表格确实构成了有帮助的知识综合。附录C提供了各查询的详细统计数据和标注者间一致性分析。

定性反馈解释了其中的原因。专家们反复指出，只需“单击一下”便能调出比较表，将大幅加快文献综述的进程，也能更轻松地比较多项研究中的工艺条件，尤其有助于前驱体筛选或确定合适的工艺窗口等任务。他们强调，这类机器可操作的表格既能支持学术研究，也能服务工业应用；例如，可以快速调查针对某种材料尝试过哪些前驱体和共反应物，或绘制报告了可接受生长速率或刻蚀速率的温度和等离子体功率范围。

与此同时，较低的实用性评分和批评意见指出了有待改进的具体方面，而非根本性的反对意见。大多数保留意见涉及：

- **缺失或不完整的参数**，例如黏附系数、表面取向和终止方式、更清晰的前驱体／共反应物区分、掺杂剂信息、每周期生长量的误差棒，或更详细的原子层沉积/原子层刻蚀周期各阶段时间参数。

- **模式和分类的改进**，包括以更一致的方式处理机理和前驱体系列，并更清楚地区分从业者不会归为“真正”原子层沉积/原子层刻蚀的工艺。

- **表格可读性**，例如避免重复行、使超长表格更加紧凑，并清晰展示关键标识信息（如材料和前驱体名称），而不是将其隐藏在晦涩难懂的资源标识符之后。

- **数据完整性与可信度**，尤其是在数据稀疏的比较中，或在表格可能通过 PDF 文件自动（部分）提取的情况下。

这些“愿望清单”问题尤其具体地展现了专家希望如何使用此类表格。例如，查询针对特定材料最常报告哪些表面取向和表面终止；按前驱体/共反应物组合对 ALD 工艺配方进行分组，并将每周期生长量随温度、等离子体功率和反应器类型的变化可视化；或者按材料、前驱体、反应物和工艺温度汇总 ALE条件。这些示例不只是简单的查找，也反映了 ALD/ALE 研究人员目前已在手动进行的多参数推理。

专家还指出了若干会影响采用意愿的因素。障碍包括：除了传统 PDF 文件外，还需额外投入精力录入机器可操作的表格；文献中的报告规范参差不齐；人们对自动提取数据的质量控制有所顾虑；以及目前缺乏关于应记录哪些工艺与性能参数的社区标准。另一方面，明确的激励因素包括：文献综述可显著节省时间、在整个社区推行模式标准化、提供便于用户编写与查询 ORKG 比较的工具，以及为日常工艺开发带来切实益处（例如，能够“一键”回答复杂的比较性问题）。

综合来看，这些发现对 RQ1 给出了肯定的回答：ALD/ALE 专家认为，基于 ORKG 的机器可操作综述既具有科学意义，也具有实际价值。他们的反馈主要指出了模式细化、数据覆盖范围和工具链方面的后续工作，而非提出概念性异议。因此，在本节其余部分中，我们将 SPARQL 得出的表格视为经专家验证的符号基线，并考察基于 PDF 和 ORKG 比较运行的神经系统在多大程度上能够逼近这一基线（RQ2 和 RQ3）。完整的领域专家调查问卷答复（包括个人评分和自由文本评论）可在我们的 Zenodo 存储库中获取 [15]。

表 31：所有实验设置下的评估结果，以宏平均相对映射相似度（RMS）的精确率、召回率和 F_1 （%）表示。设置 1 是针对 ORKG 比较的符号化 SPARQL 基线。设置 2 包括对 PDF 进行神经网络查询（2.1：专有模型；2.2a：使用单个串接 PDF 上下文的开放权重模型；2.2b：使用检索增强生成（RAG）对 PDF分块进行查询的开放权重模型）。设置 3 包括基于符号上下文增强、针对 ORKG 比较 CSV文件进行神经网络问答（3.1：专有模型；3.2：开放权重模型）。“#queries”列表示，对于每个系统，33 个 ALD+ALE 查询中有多少个产生了有效预测，可与 SPARQL 标准答案表进行评估。开放权重模型的简称：“gemma-3” = gemma-3-27b-it；“llama-3.3” = Llama-3.3-70B-Instruct；“mistral” = Mistral-Large-Instruct-2407；“qwen2.5-vl” =Qwen2.5-VL-72B-Instruct；“qwen3” = Qwen3-30B-A3B-Instruct-2507。

设置	模型类型	上下文	模型	#查询	RMS-精度 [%]	RMS-rec. [%]	RMS-F1 [%]
1	符号式	ORKG 比较（基于 SPARQL 的图）	SPARQL	33	100.0	100.0	100.0
2.1	属性	综述文章 PDF（黑箱专有 PDF 处理）	chatgpt-5.1	33	46.0	39.5	40.6
			claude-sonnet	33	61.4	63.7	61.7
			gemini-3	33	59.6	60.0	59.1
2.2a	Open	PDF文本和表格（单一拼接上下文）	gemma-3	33	25.5	25.1	24.6
			llama-3.3	33	32.6	35.9	32.8
			mistral	32	28.4	27.1	25.9
			qwen2.5-vl	33	34.1	27.6	28.2
			qwen3	32	29.1	33.9	29.7
2.2b	Open	PDF文本和表格（基于嵌入分块）	gemma-3	33	21.4	18.9	19.0
			llama-3.3	33	25.0	23.8	22.8
			mistral	33	23.8	20.9	21.1
			qwen2.5-vl	33	28.2	24.1	24.3
			qwen3	33	24.7	24.6	23.2
3.1	属性	ORKG 比较导出为 CSV（符号表格上下文）	chatgpt-5.1	33	67.8	61.5	63.0
			claude-sonnet	33	**74.0**	**76.1**	**74.2**
			Gemini-3	33	72.2	71.5	70.3
3.2	Open	ORKG 比较导出为 CSV（符号表格上下文）	gemma-3	33	44.8	47.4	44.7
			llama-3.3	33	**48.5**	**52.8**	**49.3**
			mistral	31	51.4	51.2	48.8
			qwen2.5-vl	32	46.0	37.9	39.3
			qwen3	32	40.8	46.0	40.8

4.2.2 基于 PDF 文件的神经查询：在精确科学问答任务中的性能

这里我们探讨研究问题2，即仅能查看综述文章 PDF 的大语言模型，能否很好地复现通过符号式 SPARQL 查询获得的精确表格形式答案。相应的相对映射相似度（RMS）得分（见第 4.1.3 小节）汇总于表 31，涵盖设置2.1、2.2a和2.2b。在设置 2.1 中，我们使用三种专有模型（ChatGPT-5.1、Gemini-3-Pro 和 Claude-Sonnet-4.5）的聊天界面，上传完整的 PDF，并将其内部的 PDF 处理视为黑箱。在设置 2.2a 和 2.2b 中，我们通过透明处理流程评估开放权重模型：先从每个 PDF 中提取文本和表格，再将其作为输入提供，方式是将其作为一个长上下文（2.2a），或通过对文档片段进行简单检索（2.2b）。实现细节见附录。

在这一纯基于 PDF 的范式中，设置 2.1 中的专有模型表现最佳。Claude-Sonnet-4.5 以 61.7% 的最高 RMS F_1 位居首位，其次是 Gemini-3-Pro（59.1%），ChatGPT-5.1 则为 40.6%。即使是表现最弱的专有模型，也优于设置 2.2a 和 2.2b 中的所有开放权重系统，这表明目前专有模型在 PDF 处理和文档结构处理方面，相较于我们简单的开源处理流程具有显著优势。在采用单一拼接上下文的开放权重模型（设置 2.2a）中，表现最佳的 Llama-3.3-70B-Instruct 达到 32.8% 的 RMS F_1, , 其余模型则集中在 24–30% 左右。转向检索增强版本（设置 2.2b）后，性能始终下降：五个开放权重模型的 RMS F_1 均降至 19–24% 区间，这表明我们目前采用的“检索少量文本段并据此作答”这一直接策略，还不足以可靠地完成表格类问答，甚至可能遗漏重要证据。

总体而言，这些发现表明，纯神经网络的基于 PDF 的查询虽然在一定程度上可行，但其表现仍显著低于符号化的 SPARQL 上限，并且受到 PDF 处理和上下文选择流程质量的严格制约。在这一范式下，专有系统目前树立了最高标准，但即使是其中表现最好的系统，也只能恢复约 60% 的 SPARQL RMS F_1信号。对于开放权重模型而言，差距进一步扩大，而朴素检索并不能弥合这一差距。这

促使我们进入下一小节（RQ3）：在该小节中，我们移除 PDF 瓶颈，并将相同的模型直接置于可机器操作的 ORKG 比较表中。

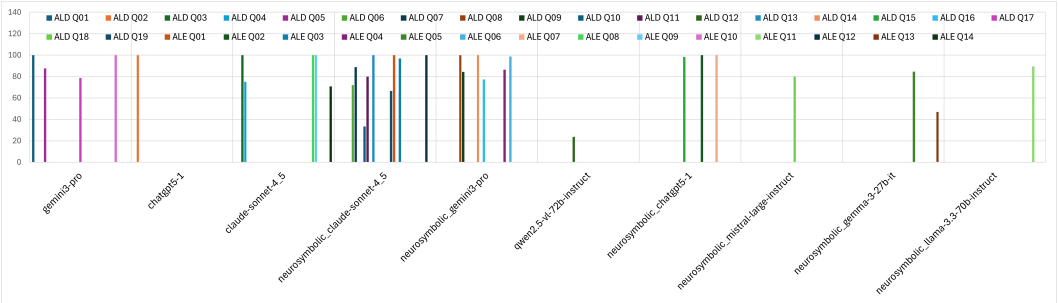
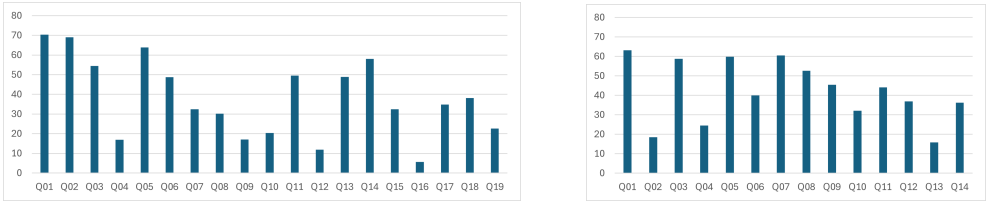
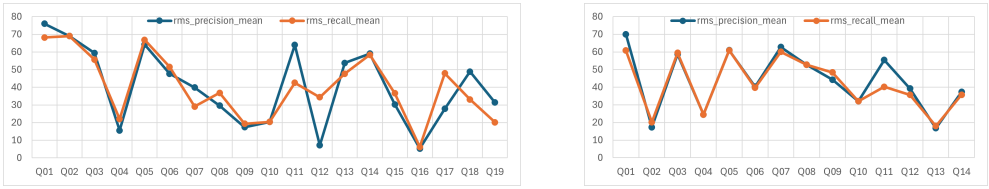


图 9：每个查询获得的最佳 RMS F_1 分数。对于每个 ALD 和 ALE 查询，我们确定与 SPARQL 黄金标准表相比具有最高 RMS F_1 分数的系统。在评估的 21 个系统中，有 10 个系统至少在一个查询上取得了最佳 RMS F_1 分数。这里，“最佳”表示在 21 个系统中针对给定查询所达到的最高 RMS F_1 分数，并不一定意味着达到 100% 的完美分数。



（a）19 个 ALD 查询的平均 RMS F_1 分数，取 21 个系统的平均值。（b）14 个 ALE 查询的平均 RMS F_1 分数，取相同 21 个系统的平均值。

图 10：ALD（左）和 ALE（右）各查询的相对映射相似度（RMS）F_1 分数。对于每个查询，我们计算每个系统的预测表格与 SPARQL 黄金标准表之间的 RMS F_1，然后对 21 个系统的 F_1 数值取平均值。较高的平均 RMS F_1 分数表示模型更容易复现符号化 SPARQL 结果的查询，而较低的分数则表示在 LLM 环境下更具挑战性的查询。



(a) 19 个 ALD 查询的平均 RMS 精确率和召回率，取 21 个神经系统和神经符号系统的平均值。(b) 14 个 ALE 查询的平均 RMS 精确率和召回率，取同样 21 个系统的平均值。

图11：ALD（左）和 ALE（右）中各查询的相对映射相似度（RMS）精确率和召回率。对于每个查询，我们计算各系统预测表格与 SPARQL 金标准表之间的 RMS 精确率和召回率，然后对 21 个系统的得分取平均值。平均精确率曲线与召回率曲线重合的点表明，系统倾向于复现金标准映射，既不添加也不遗漏条目。相反，如果两者存在明显差距，且精确率显著低于召回率，则表明存在幻觉（添加了许多 SPARQL 不支持的条目）；而精确率高于召回率则表明覆盖不足，即模型倾向于只生成金标准条目的保守子集。

4.2.3 基于 ORKG 比较的神经查询：符号化落地的优势

从 ALD/ALE 的角度来看，RQ3 提出了一个简单的问题：如果向大语言模型提供与专家所用类型相同的结构化表格（即 ORKG 比较），它们给出的答案是否会比必须自行阅读 PDF 时更接近“理想”的 SPARQL 答案？与设置 2（基于 PDF 的查询）不同，设置 3 使用相同的专有模型和开放权重模型，但将其输入限制为以 CSV 序列化的 ORKG 比较表。每个模型都会收到导出的比较表以及详细的自然语言查询。这一设置隔离了基于完全结构化、机器可操作证据进行的神经推理，使我们能够将符号上下文增强的神经查询直接与符号 SPARQL 金标准（设置 1）以及纯神经网络的基于 PDF 查询（设置 2）进行比较。所有系统的相对映射相似度（RMS）得分（见第 4.1.3 小节）汇总于表 31。

在表 31 中，最后两个主要区块（设置 3.1 和设置 3.2）清楚地展现了符号化落地带来的提升。在以 ORKG 表格为上下文时，三种专有模型的表现均显著优于仅使用 PDF 的对应模型，且最佳模型的 RMS F_1 超过了 70%。Claude-Sonnet-4.5 达到 74.2% RMS F_1, Gemini-3-Pro 为 70.3%，ChatGPT-5.1 为 63.0%。对于原子层沉积/原子层刻蚀领域的读者来说，这意味着：与提供完整 PDF 相比，若向这些系统提供整洁的工艺表，其回答中的行和数值约有三分之二到四分之三与 SPARQL 基线相符。在开放权重模型领域，Llama-3.3-70B-Instruct 仍是表现最强的模型，在设置 3.2 中上升至 49.3% RMS F_1 ，Mistral-Large-Instruct-2407 紧随其后，达到 48.8%。因此，符号化表格上下文提升了所有 模型系列的性能，并显著缩小了与 SPARQL 上限（100%）之间的差距（但并未完全弥合）。

图 9 展示了对于每个 ALD 和 ALE 查询，与 SPARQL 结果比较时，哪个系统取得了最高的 RMS F_1 。大多数查询由采用符号化落地设置 3 的系统 “ 胜出 ”，而非仅基于 PDF 的设置 2.1–2.2。换句话说，只要我们测试的系统中有模型能够较好地回答某个查询，它几乎总是在看到精选 ORKG 对比表时做到这一点，而不是在看到原始 PDF 时。这直接支持了我们的核心假设：与仅依赖大语言模型在非结构化文档上的生成能力相比，结构化机器可操作证据库更有利于精确信息与知识综合。

与此同时，符号化落地并不会让所有问题都变得容易。图 10 展示了所有 21 个系统在每个查询上的 RMS F_1 平均值。有些问题的平均得分相对较高，一旦提供 ORKG 表格，许多模型都能可靠地作答。另一些问题即使在设置 3 中仍然很难，平均 RMS F_1 有时低于 20%。这些较难的情况往往包含多个要素：严格的数值约束（例如同时限制温度、厚度或刻蚀速率）、信息分散在较长或多页表格中，以及需要按较高层级的类别（如前驱体系列、连接体类型或器件类别）进行分组或汇总，或跨多个表格联结信息。对于这类多步推理任务，筛选、分组或求平均中的细小错误很快就会导致行缺失或错误，并受到 RMS 指标的严厉惩罚。

图11通过绘制每个查询的平均 RMS 精确率和召回率，提供了互补的视角。当精确率和召回率接近时，系统往往能够复现 SPARQL 映射，而不会添加许多无根据的条目，也不会遗漏许多金标准行。精确率明显低于召回率，表明存在幻觉：模型常会输出 SPARQL 结果不支持的额外表格条目。精确率高于召回率则表明存在覆盖不足，即系统仅返回真实条目中较为保守的子集。从 PDF文件（设置 2）转向 ORKG 表格（设置 3），显著减少了幻觉和覆盖不足，但并未将其完全消除，尤其是对于上文所述较复杂的查询。

总体而言，对于 ALD/ALE 实践，RQ3 的答案是肯定的：当大语言模型以结构化的 ORKG 比较表为依据时，其表现会显著接近符号化 SPARQL 基线，并且比仅依据 PDF 文件时更加可控。与此同时，现存的错误也指出了明确的改进方向：构建更丰富、更完整的对比模式，更好地处理多约束数值筛选和跨表连接，并采用交互模式，引导模型进行系统化的筛选与聚合，而非临时生成文本。所有自然语言查询、SPARQL 表述、答案表和 ORKG 比较均已发布于 Zenodo 上的 *ALD/E* 查询数据集 （DOI: https://doi.org/10.5281/zenodo.17720429） [15]，供 ALD/ALE 社区查阅、复用和扩展。

5 讨论

我们的实验将两种截然不同的 ALD/ALE 问题回答方式并列比较，并直观呈现了各自的优势与不足。一方面，通过 SPARQL 对 ORKG 比较进行符号查询速度很快（在我们的硬件上，每次查询耗时微秒级），结果具有确定性（对相同数据执行相同查询，总会得到相同答案），且操作基于明确的材料、工艺和性能描述符模式。这与许多 JVSTA 读者对工艺数据的理解方式相符：将其视为定义清晰的表格或数据库中的行与列。
另一方面，针对 PDF 的神经网络查询表现则截然不同。当我们要求大语言模型阅读完整的综述文章 PDF 并生成表格形式的答案时，查询耗时以分钟计，而非微秒；结果可能因运行而异；模型也容易产生幻觉、误读表格以及出现基本的数值错误——即便我们为 SPARQL 问题精心设计了自然语言版本。在若干案例中，冗长的“逐步推理”链会使模型无法跟踪自身逻辑，转而给出重复或部分自相矛盾的答案，而不是逐步得出我们所需的精确对比表。

向大语言模型提供 ORKG 表格能改善情况，但无法彻底解决问题。在 ORKG 增强设置（设置 3）中，我们用导出为 CSV 的精简、符合模式的比较表替代 PDF。这显著提升了答案质量，并减少了一些幻觉：模型的表现明显接近 SPARQL 基线，也更加可预测。不过，它们仍比直接使用 SPARQL 查询慢，运行结果仍有一定差异，而且偶尔仍会误解列的语义或数值。换言之，结构化表格能让大语言模型实用得多，但无法使其成为符号查询的完美即插即用替代品。

这些观察表明，我们的主要贡献在于论证科学知识应保留一个符号层，而非声称我们构建了一个完整的神经符号系统。设置 3 最适合被理解为符号基础神经问答：ORKG 比较（一个结构化、由人工整理的表格）对语言模型加以约束和引导，而该模型本质上仍是随机的，即从生成序列中下一个可能词元的概率分布中进行采样，因此同一个问题在不同运行中可能得到略有不同的输出。我们没有改变模型架构或训练目标，也没有在网络内部集成逻辑推理模块。相反，我们通过实证表明，即便是最强的大语言模型，也不应被视为确定性科学应用中符号查询引擎的替代品。它们至多只能近似 SPARQL 的输出；若能向其提供经过整理并在 ORKG 等知识图谱中维护的符号结构，效果则会显著提升。

展望未来，我们的结果表明，在 ALD/ALE 研究中，符号问答与基于大语言模型的问答应被视为互补方式。通过 SPARQL 对 ORKG 进行符号查询，最适用于对定义明确的量进行精确、可重复的筛选和聚合（例如：“哪些适用于材料 X 的 ALD 工艺配方满足指定的温度范围和每周期生长约束？”）。在这些情况下，每一行和每一个数字都至关重要，而符号层能够提供硬性保证：对相同数据执行相同查询，始终会返回相同结果，并具有透明的溯源信息。相比之下，基于大语言模型的问答最适合处理更开放、探索性更强或需要综合分析的任务，例如解释机理、将工艺与更广泛的趋势联系起来，或将零散观察整合成连贯的叙述。这一观点与近期的 ALD 基准测试工作一致；该研究将 GPT-4 级模型评估为概念性和设计类问题的助手，而非精确复现表格的引擎 [65], 。此外，MOF-ChemUnity 将金属有机框架的文献衍生知识库与由大语言模型驱动的检索和结构–性质推理辅助相结合 [53]。在这两种情况下，只有当大语言模型建立在经过整理的知识之上，而非成为定量真值的唯一来源时，它们才最具说服力。

在符号层面，我们的工作也与材料科学领域由文献构建的大型知识图谱相关，例如 MatKG [61], ，它能够从数百万篇材料论文中自动提取实体和关系，并将其作为可通过 SPARQL 查询的 RDF/CSV 知识图谱呈现。MatKG 强调广泛的自动化挖掘覆盖范围，而我们基于 ORKG 的比较则聚焦于数量较少的综述表格，这些表格经过明确的模式设计和专家整理，并针对原子层沉积/原子层刻蚀提供流程级细节。这些工作共同展现了符号基础设施的不同层次：从

范围广泛、自动生成的图，到高保真、机器可操作的综述表格；后者可作为精确查询和未来神经符号集成的金标准目标。

鉴于本文阐述了符号知识的价值，自然的下一步是探索符号知识与神经模型更紧密的集成。一种方向是直接使用符号化资料——比较表、ORKG 三元组以及 SPARQL 查询—结果对——训练或微调大语言模型，使其更擅长精确的表格类操作，同时仍能支持趋势外推和假设生成（例如，建议候选前驱体、工艺窗口或后续实验）。然而，要使这些方法有效，底层符号数据必须经过验证 且具备大规模。由于当前大语言模型已经在海量自然语言文本上完成训练，瓶颈便转向构建足够丰富的金标准符号数据集。这需要通过工作流程，以极高的准确率（理想情况下，表格重构和关键字段提取的准确率远高于 95%）从大量论文中挖掘机器可操作的知识，并将自动提取与有针对性的人工验证相结合。因此，展望未来，我们邀请 ALD/ALE 社区发布综述表格的机器可操作版本，并与传统 PDF 一并提供，作为不断扩充的金标准集合的一部分（例如，存入 ORKG 和关联存储库）。另一个互补方向是设计混合系统：由确定性符号核心（例如，基于 ORKG 的 SPARQL 引擎）负责精确筛选和聚合，同时由神经组件协助组织自然语言查询、对结果进行排序和聚类，并解释其中的模式。在这些情形中，我们的研究结果指向 ALD/ALE 工作流程的一项明确设计原则：可靠的定量查询应以 ORKG 等符号基础为锚点，而神经模型则作为强大但近似的辅助层发挥作用，适用于探索、解释和创意生成，但不应被视为工艺关键数值的独立真值来源。

6 结论

本研究展示了如何将 ALD/原子层刻蚀综述知识从静态 PDF 表格转化为机器可操作层，从而支持精准、可重复的查询。我们从近期的九篇综述文章中，将 18 个涵盖多项研究的工艺表格转化为 ORKG 比较，并制定了 33 个自然语言问题及相应的 SPARQL 查询，将其作为 ALD/E 查询数据集 [15]发布。一项针对三位 ALD/原子层刻蚀专家开展的调查证实，这些机器可操作的综述具有科学意义，并在前驱体筛选、工艺窗口比较和机理概述等任务中具有实际用途；调查还指出了完善模式、补充缺失参数的具体机会。对比三种回答相同问题的方式——通过 SPARQL 对 ORKG 进行符号查询、对综述 PDF 进行纯神经网络查询，以及以 ORKG 比较表为依据的神经网络查询——可以看出各自的作用分工十分明确：通过 SPARQL 查询 ORKG 仍是获得精确、可复现答案及透明溯源信息的金标准；大语言模型则最适合作为其上的辅助层，尤其是在以结构化 ORKG 表格为依据时，可用于组织问题、探索模式和生成假设。

展望未来，我们的研究结果表明，应加强并深化符号工具与神经工具之间的联系。在符号方法方面，我们认为需要建立社区工作流程，定期发布综述表格的机器可操作版本，并与 PDF 文件一同提供，从而创建规模庞大且经过验证的语料库，既可作为黄金标准资源，也可用作未来神经符号模型的训练材料。在神经方法方面，由确定性符号核心负责精确筛选与聚合，同时由大语言模型处理查询构建、结果探索和解释的混合系统，尤其值得期待。最后，本研究也补充了新兴的多模态科学人工智能研究；在这类研究中，模型不仅需要解读文本和表格，还需要解读图表，因为图表承载了原子层沉积/原子层刻蚀及相关领域中大量的定量信息 [14]。综上，这些方向勾勒出一条发展路径，旨在打造科学助手，使其能够可靠地回答精确问题、帮助理解复杂数据集，并最终以整合的方式跨文本、表格和图像进行推理。

致谢与经费披露

使用五个开放权重大语言模型开展的实验得到了 KISSKI 高性能计算基础设施的支持， [13], 尤其是通过 ChatAI API 服务（https://kisski.gwdg.de/en/leistungen/ 2-02-llm-service/）。该集中式服务提供对拥有数十亿参数的超大型语言模型的访问

这些模型的计算需求通常超出学术研究环境中本地图形处理器资源的承载能力，从而减轻了重复部署以及耗时的临时查询设置的需求。

本研究由“面向可持续半导体工艺与制造技术的人工智能感知路径”项目资助，该项目由英特尔公司与德国达姆施塔特的默克集团共同赞助。

数据可用性声明

本文配套代码已发布于 Github：https://github.com/sciknoworg/ ald-ale-orkg-review，并附有详细文档。本研究生成并研究的数据集已发布于 Zenodo [15]，可通过此 DOI 访问：https://doi.org/10.5281/ zenodo.17720429。

参考文献

[1] Amr Ghazy、David Zanders、Anjana Devi 和 Maarit Karppinen。稀土掺杂剂在原子层沉积薄膜中的应用：主体材料、掺杂剂与应用，2025年。网址：https://orkg.org/ comparison/R1469955。[2] Amr Ghazy、David Zanders、Anjana Devi 和 Maarit Karppinen。稀土基薄膜的原子层沉积工艺，2025年。网址：https://orkg.org/comparison/ R1471077。[3] Amr Ghazy、David Zanders、Anjana Devi 和 Maarit Karppinen。稀土-有机杂化薄膜的稀土原子层沉积/分子层沉积工艺，2025年。网址：https://orkg.org/comparison/R1470110。[4] Andreas Fischer、Aaron Routzahn、Steven M. George 和 Thorsten Lill。按材料和反应物汇总已报道的热原子层刻蚀工艺，2025年。网址：https: //orkg.org/comparison/R1563034。[5] Anthropic。Claude Sonnet 4.5 发布介绍。https://www.anthropic.com/news/ claude-sonnet-4-5，2025年9月。访问日期：2025年11月21日。[6] Sören Auer、Allard Oelen、Muhammad Haris、Markus Stocker、Jennifer D’Souza、Kheir Eddine Farfar、Lars Vogt、Manuel Prinz、Vitalis Wiens 和 Mohamad Yaser Jaradeh。利用知识图谱提升科学文献的可及性。《图书馆、研究与实践》，44(3):516–529，2020年。[7]Sören Auer、Jennifer D’Souza、Kheir Eddine Farfar、Mohamad Yaser Jaradeh、Azanzi Jiomekong、Oliver Karras、Allard Oelen、Lauren Snyder、Markus Stocker 和 Lars Vogt。开放研究知识图谱：一种大规模神经符号知识组织系统。载于《神经符号人工智能与知识图谱手册》，第385–420页。IOS出版社，2025年。[8] Caio Viktor S Avila、Vânia MP Vidal、Wellington Franco 和 Marco A Casanova。基于 ChatGPT 的文本到 SPARQL 实验。载于*2024*年第*18*届*IEEE*国际语义计算会议（*ICSC*），第277–284页。IEEE，2024年。[9] Vannevar Bush 等。《诚如所思》。《大西洋月刊》，176(1):101–108，1945年。[10] Chang Fang、Yanqiang Cao、Di Wu 和 Aidong Li。代表性材料的典型热原子层刻蚀（ALE）工艺，2025年。网址：https://orkg.org/comparison/ R1560222。[11] Marco Cremaschi、Federico Belotti、Jennifer D’Souza 和 Matteo Palmonari。Mammotab 25：用于语义表格解释的大规模数据集——训练、测试与弱点检测。载于国际语义网会议，第131–148页。施普林格，2025年。[12] Doo San Kim、Ju Eun Kim、You Jung Gill、Yun Jong Jang、Ye Eun Kim、Kyong Nam Kim、Geun Young Yeom 和 Dong Woo Kim。已报道的金属各向异性和各向同性原子层刻蚀（ALE）工艺综述，2025年。网址：https://orkg.org/comparison/R1563131。

[13] Ali Doosthosseini、Jonathan Decker、Hendrik Nolte 和 Julian Kunkel。SAIA：面向高性能计算服务的无缝 Slurm 原生解决方案，2025年7月。国际标准连续出版物号 2693-5015。网址：https: //www.researchsquare.com/article/rs-6648693/v1。[14] Jennifer D'Souza。迈向通用科学人工智能：科学图像的多模态理解，2025年9月。网址：https://doi.org/10.5281/zenodo.17130928。[15] Jennifer D'Souza、Eleni Poupaki、Alex Watkins 和 Randall Higuchi。ALD/E 神经符号查询基准：针对机器可操作的 ORKG 比较的33个科学查询，2025年11月。网址：https://doi.org/10.5281/zenodo.17720429。[16] Lisa Ehrlinger 和 Wolfram Wöß。探讨知识图谱的定义。*SEMANTiCS*（海报、演示、*SuCCESS*），48(1-4):2，2016年。[17] Vincent Emonet、Jerven Bolleman、Severine Duvaud、Tarcisio Mendes de Farias 和 Ana Clau- dia Sima。基于大语言模型的联邦知识图谱自然语言到 SPARQL 查询生成，2024年。网址：https://arxiv.org/abs/2410.06062。[18] Wenqi Fan、Pangjing Wu、Yujuan Ding、Liangbo Ning、Shijie Wang 和 Qing Li。迈向检索增强型大语言模型：数据管理与系统设计。载于*2025* 年第 *41* 届 *IEEE* 国际数据工程会议（*ICDE*），第4509–4512页。IEEE，2025年。[19] Chang Fang、Yanqiang Cao、Di Wu 和 Aidong Li。热原子层刻蚀：机理、材料与前景。《自然科学进展：国际材料》，28(6):667–675，2018年。[20] Andreas Fischer、Aaron Routzahn、Steven M George 和 Thorsten Lill。热原子层刻蚀：综述。《真空科学与技术*A*辑》，39(3)，2021年。[21] G. S. Oehrlein、Donald Metzler 和 C. Li。原子层刻蚀（ALE）所用材料与化学体系综述，2025年。网址：https://orkg.org/comparison/R1562672。[22] Steven M George。原子层沉积：综述。《化学评论》，110(1):111–131，2010年。[23] Steven M George。热原子层刻蚀的机理。《化学研究述评》，53(6):1151–1160，2020年。[24] Amr Ghazy、David Zanders、Anjana Devi 和 Maarit Karppinen。采用原子层沉积制备的 Er^{3+}基 MOSLED 器件的性能比较，2025年。网址：https://orkg.org/ comparisons/R1469991。[25] Amr Ghazy、David Zanders、Anjana Devi 和 Maarit Karppinen。基于稀土元素的功能薄膜原子层沉积和分子层沉积。《先进材料界面》，12(4):2400274，2025年。[26] Aidan Hogan、Eva Blomqvist、Michael Cochez、Claudia d'Amato、Gerard De Melo、Claudio Gutierrez、Sabrina Kirrane、José Emilio Labra Gayo、Roberto Navigli、Sebastian Neumaier 等。知识图谱。《*ACM* 计算调查》（*Csur*），54(4):1–37，2021年。[27] Daeun Hong、Yongjae Kim 和 Heeyeop Chae。面向纳米级半导体器件的二氧化硅原子层刻蚀：综述。《应用科学与融合技术》，33(1):1–6，2024年。[28] Daeun Hong、Yongjae Kim 和 Heeyeop Chae。氟碳前驱体在硅各向异性原子层刻蚀中的应用。网址：https://orkg.org/comparisons/R1560949。[29] Daeun Hong、Yongjae Kim 和 Heeyeop Chae。硅各向异性原子层刻蚀中的离子能量窗口。网址：https://orkg.org/comparisons/R1560977。[30] Daeun Hong、Yongjae Kim 和 Heeyeop Chae。硅各向异性原子层刻蚀中的刻蚀速率选择性。网址：https://orkg.org/comparisons/R1561025。

[31] Daeun Hong、Yongjae Kim 和 Heeyeop Chae。腔壁条件对 SiO 各向异性原子层刻蚀的影响。网址：https://orkg.org/comparisons/R1561023。[32] Daeun Hong、Yongjae Kim 和 Heeyeop Chae。SiO 的热诱导各向同性原子层刻蚀机制。2025年。网址：https://orkg.org/comparisons/R1561046。[33] Daeun Hong、Yongjae Kim 和 Heeyeop Chae。SiO 的等离子体辅助各向同性原子层刻蚀机制。网址：https://orkg.org/comparisons/R1562478。[34] Keren J Kanarik、Thorsten Lill、Eric A Hudson、Saravanapriyan Sriraman、Samantha Tan、Jeffrey Marks、Vahid Vahedi 和 Richard A Gottscho。半导体行业原子层刻蚀综述。《真空科学与技术*A*辑》，33(2)，2015年。[35] Keren J Kanarik、Samantha Tan 和 Richard A Gottscho。原子层刻蚀：重新思考刻蚀艺术。《物理化学快报》，9(16):4814–4821，2018年。[36] Erwin Kessels、Anjana Devi、Jin-Seong Park、Mikko Ritala、Angel Yanguas-Gil 和 Claudia Wiemer。原子层沉积。*NatureReviews Methods Primers*，5(1):66，2025年。[37] Markku Leskelä 和 Mikko Ritala。原子层沉积（ALD）：从前驱体到薄膜结构。《薄固体薄膜》，409(1):138–146，2002年。[38] Fangyu Liu、Julian Eisenschlos、Francesco Piccinno、Syrine Krichene、Chenxi Pang、Kenton Lee、Mandar Joshi、Wenhu Chen、Nigel Collier 和 Yasemin Altun。DePlot：通过图表到表格转换进行单次视觉语言推理。载于 Anna Rogers、Jordan Boyd-Graber 和 Naoaki Okazaki 编，计算语言学协会年会（*ACL 2023*）论文集，第10381–10399页，加拿大多伦多，2023年7月。计算语言学协会。doi：10.18653/v1/2023.findings-acl.660。网址：https://aclanthology.org/2023.findings-acl.660/.[39] 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*2021*年*IEEE*冬季计算机视觉应用会议（*WACV*）。计算机视觉基金会，2021年1月。网址：
h
t
t
p
s://
w
w
w.
m
i
c
r
o
s
o
f
t.
c
o
m/
en-us/research/publication/chartocr-data-extraction-from-charts-images-via-a-deep-hybrid-framework/。[40] Ahmed Masry、Do Xuan Long、Jia Qing Tan、Shafiq Joty 和 Enamul Hoque。ChartQA：一项关于图表问答的基准测试，结合视觉推理和逻辑推理。载于 Smaranda Muresan、Preslav Nakov 和 Aline Villavicencio 编，计算语言学协会年会（*ACL 2022*）论文集，第2263–2279页，爱尔兰都柏林，2022年5月。计算语言学协会。doi：10.18653/v1/2022.findings-acl.177。网址：https://aclanthology.org/2022.findings-acl.177/。[41] Ville Miikkulainen、Markku Leskelä、Mikko Ritala 和 Riikka L Puurunen。原子层沉积制备无机薄膜的结晶度：综述与总体趋势。《应用物理学杂志》，113(2)，2013年。[42] Natasha Noy、Yuqing Gao、Anshu Jain、Anant Narayanan、Alan Patterson 和 Jamie Taylor。工业级知识图谱：经验与挑战——五家各具特色的科技公司展示实践方法。《*Queue*》，17(2):48–75，2019年。[43] GS Oehrlein、D Metzler 和 C Li。原子层刻蚀的转折点：综述。《*ECS*固态科学与技术杂志》，4(6):N5041，2015年。[44] Allard Oelen、Mohamad Yaser Jaradeh、Markus Stocker 和 Sören Auer。利用学术知识图谱生成 FAIR 文献综述。载于*2020*年*ACM/IEEE*数字图书馆联合会议论文集，第97–106页，2020年。[45] Allard Oelen、Markus Stocker 和 Sören Auer。从综述文章表格构建学术知识图谱。载于 Emi Ishita、Natalie Lee San Pang 和 Lihong Zhou 编，《社会巨变时期的数字图书馆》，第373–389页，Cham，2020年。施普林格国际出版公司。

[46] Allard Oelen、Markus Stocker 和 Sören Auer。SmartReviews：构建面向人类和机器的综述文章表征。载于 Hao-Ren Ke、Chei Sian Lee 和 Kazunari Sugiyama 编，《迈向开放可信的数字社会》，第105–114页，Cham，2021年。施普林格国际出版公司。[47] OpenAI。GPT-5.1：更智能、更善于对话的 ChatGPT。https://openai.com/index/gpt-5-1/，2025年11月。访问日期：2025年11月21日。[48] Peter Michael Piechulla、Mingliang Chen、Aristeidis Goulas、Riikka Puurunen 和 J. Ruud van Ommen。荧光粉材料上原子层沉积涂层的比较研究，2025年。网址：https://orkg.org/comparison/R1469383。[49] Peter Michael Piechulla、Mingliang Chen、Aristeidis Goulas、Riikka Puurunen 和 J. Ruud van Ommen。药用颗粒材料上原子层沉积涂层的比较研究，2025年。网址：https://orkg.org/comparison/R1469594。[50] Sundar Pichai、Demis Hassabis 和 Koray Kavukcuoglu。Gemini 3。https://blog.google/products/gemini/gemini-3/，2025年11月。访问日期：2025年11月21日。[51] Peter M Piechulla、Mingliang Chen、Aristeidis Goulas、Riikka L Puurunen 和 J. Ruud van Ommen。1988至2023年颗粒材料上的原子层沉积：技术、材料与应用的定量综述。*arXiv*预印本 *arXiv:2506.17725*，2025年。[52] Georgi Popov、Miika Mattinen、Anton Vihervaara 和 Markku Leskelä。热原子层沉积化学的最新趋势。《真空科学与技术*A*辑》，43(3)，2025年。[53] Thomas Michael Pruyn、Amro Aswad、Sartaaj Takrim Khan、Ju Huang、Robert Black 和 Seyed Mohamad Moosavi。MOF-ChemUnity：以文献为依据、用于金属有机框架研究的大语言模型。《美国化学会志》，2025年。[54] Riikka L Puurunen。原子层沉积表面化学：三甲基铝/水工艺案例研究。《应用物理学杂志》，97(12)，2005年。[55] Doo San Kim、Ju Eun Kim、You Jung Gill、Yun Jong Jang、Ye Eun Kim、Kyong Nam Kim、Geun Young Yeom 和 Dong Woo Kim。金属的各向异性/各向同性原子层刻蚀。《应用科学与融合技术》，29(3):41–49，2020年。[56] Xia Sang 和 Jane P Chang。定向原子层刻蚀中的物理和化学效应。《物理学杂志*D*辑：应用物理》，53(18):183001，2020年。[57] Amit Singhal。知识图谱简介：实体，而非字符串。https://blog.google/products/search/introducing-knowledge-graph-things-not/，2012年5月。谷歌官方博客，访问日期：2025年11月15日。[58] Amit Singhal 等。现代信息检索：简要综述。*IEEE*数据工程通报，24(4):35–43，2001年。[59] Tuomo Suntola。原子层外延。《材料科学报告》，4(5):261–312，1989年。[60] J Ruud van Ommen、Aristeidis Goulas 和 Riikka L Puurunen。原子层沉积。《柯克-奥思默化学技术百科全书》，第1–42页，2021年。[61] Vineeth Venugopal 和 Elsa Olivetti。MatKG：材料科学领域自主生成的知识图谱。《科学数据》，11(1):217，2024年。[62] Mark D Wilkinson、Michel Dumontier、IJsbrand Jan Aalbersberg、Gabrielle Appleton、Myles Axton、Arie Baak、Niklas Blomberg、Jan-Willem Boiten、Luiz Bonino da Silva Santos、Philip E Bourne 等。科学数据管理与维护的 FAIR指导原则。《科学数据》，3(1):1–9，2016年。[63] Xiao Sang 和 Jane P Chang。半导体材料已报道的精选原子层刻蚀（ALE）工艺，2025年。网址：https://orkg.org/comparison/R1560825。

[64] Derong Xu、Xinhang Li、Ziheng Zhang、Zhenxi Lin、Zhihong Zhu、Zhi Zheng、Xian Wu、Xiangyu Zhao、Tong Xu 和 Enhong Chen。通过自适应多方面检索增强，利用大语言模型进行知识图谱问答。载于*AAAI*人工智能大会论文集，第39卷，第25570–25578页，2025年。[65] Angel Yanguas-Gil、Matthew T Dearing、Jeffrey W Elam、Jessica C Jones、Sungjoon Kim、Adnan Mohammad、Chi Thang Nguyen 和 Bratin Sengupta。大语言模型在材料合成中的基准测试：以原子层沉积为例。《真空科学与技术*A*辑》，43(3)，2025年。[66] Jihong Yim、Oili ME Ylivaara、Markku Ylilammi、Virpi Korpelainen、Eero Haimi、Emma Verkama、Mikko Utriainen 和 Riikka L Puurunen。基于饱和曲线的原子层沉积共形性分析：横向高深宽比通道中的氧化铝。《物理化学化学物理》，22(40):23107–23120，2020年。[67] Jihong Yim、Oili M. E. Ylivaara、Markku Ylilammi、Virpi Korpelainen、Eero Haimi、Emma Verkama、Mikko Utriainen 和 Riikka L. Puurunen。高深宽比结构中原子层沉积氧化铝前驱体黏附系数的比较，2025年。网址：https://orkg.org/ comparisons/R1469158。[68] Zhimin Yu、Kang Yuan、Yang Yang 和 Jiaming Sun。低温下在硅上通过原子层沉积制备多晶铒掺杂$Y_3Ga_5O_{12}$纳米薄膜及其电致发光性能探索。《纳米尺度》，14(29):10540–10548，2022年。[69] Wayne Xin Zhao、Kun Zhou、Junyi Li、Tianyi Tang、Xiaolei Wang、Yupeng Hou、Yingqian Min、Beichen Zhang、Junjie Zhang、Zican Dong、Yifan Du、Chen Yang、Yushuo Chen、Zhipeng Chen、Jinhao Jiang、Ruiyang Ren、Yifan Li、Xinyu Tang、Zikang Liu、Peiyu Liu、Jian-Yun Nie 和 Ji-Rong Wen。大语言模型综述，2025年。网址：https://arxiv.org/abs/ 2303.18223。

附录

A 实验设置2 – 详细技术说明

在此设置中，我们将每篇综述文章的 PDF 视为唯一证据来源，以评估神经方法的性能上限。专有语言模型和开放权重语言模型都会收到 SPARQL 查询的详细自然语言等价形式，以及文章 PDF（或从中提取的文本和表格），但不会 直接访问 ORKG 表示。因此，模型必须在单个神经推理步骤中自行解析 PDF、定位相关段落和表格，并汇总所需数据。将其输出与设置 1 中的 SPARQL 金标准进行比较，我们便可以量化：对于精确的表格型科学问题，纯神经的基于 PDF 的查询能够在多大程度上接近符号方法性能。

A.1 专有语言模型（神经）查询。 在此，我们使用三种最先进的对话式生成式人工智能模型的专有聊天界面——ChatGPT-5.1（OpenAI，发布日期：2025年11月12日）[47], Gemini-3-Pro（Google，发布日期：2025年11月18日）[50], 和 Claude-Sonnet-4.5（Anthropic，发布日期：2025年9月29日）[5]。截至本文提交之日（2025年12月13日），这些系统代表了商用托管大语言模型中能力最强的几种。我们向每个模型提供 SPARQL 查询的详细自然语言等价形式以及综述文章的 PDF。此设置旨在检验专有模型作为黑箱系统、在 PDF 处理工作流未公开的情况下，能否仅凭 PDF 直接回答精确的科学问题。然而，专有模型提供的实验控制有限：其内部 PDF 解析流程不透明，无法检查或修改其上下文整合策略，而且其输出无法复现或系统性调试。因此，为了实现全面透明、可复现性和精细的实验控制，我们还采用了开放权重模型。

A.2 开放权重语言模型（神经网络）查询。此设置与设置 2 类似，但我们评估的不是专有模型，而是性能出色的开放权重大语言模型。我们使用 pdfplumber 从每份 PDF 中提取文本和表格内容，并将这些内容作为

输入提供给模型。我们评估了五种开放权重模型：gemma-3-27b-it、llama-3.3-70B-Instruct、qwen2.5-vl-72b-instruct、qwen3-30b-a3b-instruct-2507 和 mistral-large-instruct。这些模型因在大规模表格推理基准 [11] 中展现出强大实力而入选。

a) 上下文注入提示（单轮文档问答）。在此变体中，使用 pdf_extractor.PDFExtractor 提取每篇论文的全文和所有表格，并将其拼接成一个长上下文窗口。随后，将这份完整的文章表示与详细的自然语言查询一同直接传递给开放权重模型，具体实现见 context_injected_doc_qa.py。模型必须仅依据未经筛选的全文档上下文作答。虽然这种方法无法控制噪声或冗余，但它能确保表格保持完整，且所有证据都可供模型使用。

b) 基于 *PDF* 片段的检索增强生成（*RAG*）。与设置 3.1 中一样，我们首先使用 PDFExtractor 提取所有文本和表格。我们不会将整个文档提供给模型，而是将提取出的文本划分为连续的文本段，每段约 8 000 个字符。在我们的语料库中，每个文本段通常对应一至三页 PDF 页面，长度足以包含大多数表格，但某些较宽或跨多页的表格仍可能被拆分。然后，使用本地 Ollama 编码器（默认：nomic-embed-text:latest）将每个文本段转换为稠密数值向量（即“嵌入”），从而为每篇论文生成一个向量索引，具体实现见 rag_pdf_segments_doc_qa.py。

在查询时，使用同一个编码器对详细的自然语言查询进行嵌入，并计算其与所有片段嵌入的余弦相似度 [58]。我们检索排名最高的单个文本段（TOP_K = 1），并按索引顺序扩展选择范围，将相邻文本段也纳入其中，以便重建可能跨越文本段边界的表格内容。我们仅将这些检索片段和详细查询提供给模型。

潜在局限。 由于检索增强生成引入了明确的检索阶段，其表现不仅取决于语言模型，也取决于我们如何为每篇论文建立索引，以及如何针对给定查询选择文本段。在我们的实现中，这些设计选择会带来若干潜在的失效模式：

- 分段效应。 PDFExtractor 格式化后的 PDF 输出被切分为连续的 8k 字符文本段，大致相当于几页文本和表格。我们不在文本段之间设置重叠（CHUNK_OVERLAP = 0），以避免在不同文本块中重复表格行，但这也意味着大型综述表格可能恰好在文本段边界处被截断。尽管检索步骤总会将检索到的排名最高的文本段向两侧各扩展两个相邻文本段，但跨越这一局部范围的表格，或是在论文后文以“续表”形式再次出现的表格，在模型的上下文中仍可能只呈现部分内容。

- 嵌入敏感性。稠密检索依赖嵌入模型（nomic-embed-text），该模型会将每个文本段和每条查询映射为高维数值向量，使语义相似的文本彼此靠近。因此，检索质量取决于这个通用编码器对专业原子层沉积／原子层刻蚀术语，以及工艺变体、前驱体化学或温度范围之间细微差异的捕捉能力。如果嵌入空间未能清晰地区分这些概念，那么包含相关表格行的文本段可能会获得低于相关性较低文本段的相似度分数。

- *Top-k* 检索与局部窗口化。对于每条查询，我们只检索相似度最高的单个文本段（TOP_K = 1），然后以此“锚点”为中心，加入紧邻的文本段。这种做法有利于构建紧凑的上下文窗口，并使检索增强生成设置与全上下文设置 3.1 保持可比；但它也假定模型所需的大部分证据都集中在论文中一个连续的区域内。隐含涉及多个相距较远区域的查询（例如，表格中跨页拆分的所有 SiO_2 工艺，或表格与文本其他位置的说明性段落）可能需要分散在非相邻文本段中的信息，而这种单锚点策略无法将这些信息一并检索出来。

- 全局文档上下文丢失。与设置 3.1 中模型一次性查看提取出的整篇论文不同，检索增强生成变体刻意将模型限制在文档的一小段内容内，并要求它

仅根据检索片段作答。对于需要汇总所有行、比较多种材料或交叉核对文章不同部分所述约束的面向表格的查询，这个狭窄的窗口可能会遗漏恰好位于检索邻近范围之外的相关行或条件。在这种情况下，检索阶段可能会移除完整上下文设置原本能够提供的关键上下文。

因此，这一设置检验了一种检索增强方法，旨在缩短上下文长度并突出相关证据，同时也引入了依赖于检索和分段的行为，这些行为可能会影响下游性能。

B 神经问答表格结果的评估指标

为评估研究问题2和 RQ3，我们逐查询衡量神经系统生成的表格与 SPARQL 标准答案表的匹配程度。直观而言，我们希望采用一种指标，它（i）不受行和列排列顺序的影响，（ii）能够区分轻微的数值或文本偏差与完全错误的条目，以及（iii）同时体现精确率（多余的行或单元格）和召回率（缺失的行或单元格）。我们采用 Liu 等人 [38] 作为 Google Research 工作的一部分所提出的相对映射相似度（RMS）指标，该指标专为图表和表格问答系统的表格级评估而设计。我们基于官方 DePlot 指标实现了 RMS 评估，代码见
https://github.com/sciknoworg/ald-ale-orkg-review/blob/main/eval_deplot_rms/README.md。此实现说明了我们如何将 Google 的 DePlot 代码（https://github.com/google-research/google-research/tree/master/deplot）适配至 ALD/ALE SPARQL 和神经查询结果数据集；该数据集以 Zenodo 转储文件 [15]的形式提供。

按照 Liu 等人的做法，RMS 建立在早期的相对数字集相似度（RNSS）指标[39, 40], 之上；该指标仅比较预测表格与黄金表格中的数值无序集合。令 $P = \{p_i\}_{i=1}^{N}$ 和 $T = \{t_j\}_{j=1}^{M}$ 分别表示预测数字集和目标数字集。两个数字之间的相对数值误差为

$$D(p, t) = \min\left(1, \frac{|p - t|}{|t|}\right). \tag{1}$$

Give在成对距离 D的 $N \times M$ 矩阵（p_i，t_j）中，找到一个最小成本的二分匹配找到$X \in \{0,1\}^{N \times M}$。RNSS 随后定义为

$$\mathrm{RNSS} = 1 - \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} X_{ij}\, D(p_i, t_j)}{\max(N, M)}. \tag{2}$$

虽然 RNSS 可用于数值合理性检查，但它忽略了表格中数值的位置（即它们所属的行和列），舍弃所有非数值内容，也不区分精确率和召回率。

RMS 则将表格视为一个无序集合，其中包含从行标题和列标题到单元格值的映射。预测表格中的每个条目表示为 $p_i = (p_i^r,\ p_i^c,\ p_i^v)$，其中 p_i^r和 p_i^c分别是行标题和列标题，p_i^v是单元格值；黄金表格则以类似方式表示为$t_j = (t_j^r,\ t_j^c,\ t_j^v)$。为了比较两个条目 p_i和 t_j，RMS 综合考虑 (i) 拼接后的键 $p_i^r \| p_i^c$和 $t_j^r \| t_j^c$之间的文本相似度，以及 (ii) 单元格值 p_i^v和 t_j^v之间的相似度。

文本相似度通过归一化莱文斯坦距离 $NL_\tau(\cdot,\ \cdot)$ 衡量，阈值为 τ，该阈值会将较大的编辑距离截断为 1：

$$s_{\text{key}}(i, j) = 1 - NL_\tau\big(p_i^r \| p_i^c,\ t_j^r \| t_j^c\big), \tag{3}$$

使得 $s_{\text{key}}(i,\ j) \in [0, 1]$。数值使用相对误差和第二个阈值 θ进行比较；相应的相似度为

$$s_{\text{val}}(i, j) = 1 - \min\left(1, \frac{|p_i^v - t_j^v|}{|t_j^v|}\right), \tag{4}$$

对于差异极大的情况，再次截断为 0。当数值不是数字时， s_{val} 会回退到使用 NL_τ 的文本相似度项。两个映射之间的总体相似度为

$$s(i,j) = s_{\text{key}}(i,j)\, s_{\text{val}}(i,j), \tag{5}$$

只有当键（行+列）和对应的值都匹配良好时，该值才会接近 1。

为了比较预测表格 $P = \{p_i\}_{i=1}^{N}$ 与黄金表格 $T = \{t_j\}_{j=1}^{M}$，RMS 首先根据键相似度构建一个 $N \times M$ 代价矩阵，并求解最小代价二分图匹配问题，以获得二元指派矩阵 $X \in \{0,1\}^{N\times M}$。基于该匹配，RMS 精确率和召回率定义如下：

$$\text{RMS}_{\text{precision}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} X_{ij}\, s(i,j), \tag{6a}$$

$$\text{RMS}_{\text{recall}} = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M} X_{ij}\, s(i,j), \tag{6b}$$

$$\text{RMS}_{\text{F1}} = \frac{2\cdot \text{RMS}_{\text{precision}}\cdot \text{RMS}_{\text{recall}}}{\text{RMS}_{\text{precision}} + \text{RMS}_{\text{recall}}}. \tag{6c}$$

由于 RMS 将表格视为无序的（行、列、值）映射集合，并同时考虑原始布局和转置布局，因此其设计本身便具有对行和列排列以及表格转置的不变性。

在我们的实验中，对于每个查询，我们将 SPARQL 结果表作为黄金表格 T，并将各神经系统（仅神经系统和符号上下文增强神经系统设置）的输出作为预测表格 P。我们使用 DePlot 原始实现的默认阈值（针对文本和数值相似度），通过我们的评估脚本计算 RMS 精确率、召回率和 F1 分数。我们报告每个系统和评估设置下逐查询的 RMS 得分，以及跨查询的宏平均 RMS，从而量化黄金表格中有多少行和单元格被恢复（召回率），以及预测条目中有多少确实对应于 SPARQL 派生证据（精确率）。

C 结果——领域专家印象的详细分析（RQ1）

此处我们详细介绍 RQ1 的专家评估结果。结果基于 3 位领域专家填写的调查问卷，具体如下。

有意义性。对于全部 33 个查询，专家均认为问题本身在科学上定义明确。有意义性平均评分为 $4.54/5$，其中 96% 的评分为 ≥ 4，且 33 个查询中有 32 个的有意义性平均评分至少为 4。ALE 查询的评分略高（平均值为 4.69），高于 ALD 查询（平均值为 4.42），且所有 ALE 评分均在 4–5 之间。如图 7 所示（另见图 7a 和 7b 中按领域划分的面板），各查询、各标注者的评分点紧密聚集在量表的上部，只有少数孤立评分低于 3（例如 ALD Q11）。这表明，整体而言，该查询集与专家对 ALD 和 ALE 中合理关系、恰当概念及相关用例的认知相符。

在 ALD 查询中，有六条查询的平均分达到满分 5.00（Q09、Q10、Q13、Q14、Q17、Q18），评分最低的是 Q11，平均分为 3.67——这是两个领域中唯一平均分低于 4.00 的查询。在 ALE 方面，有四条查询获得满分 5.00（Q02、Q05、Q07、Q08），评分最低的几条查询（Q10、Q12、Q13）平均分也达到 4.33。在单个标注者层面，ALD 的平均评分范围为 4.21 至 5.00，ALE 则为 4.29 至 4.93，反映出不同评估者在严格程度上存在中等程度的差异。

通过 Fleiss' κ 和 Krippendorff's α_{ordinal}测得的正式标注者间一致性较低（ALD：$\kappa = 0.021$，$\alpha_{\text{ordinal}} = 0.081$；ALE：$\kappa = -0.114$，$\alpha_{\text{ordinal}} = -0.088$）。当评分明显偏向量表的一端时，这是一种已知的统计假象：当几乎所有分数都是 4 或 5 时，即便只是轻微分歧（例如 4 与 5），也可能导致 κ 数值偏低或为负，尽管总体上存在广泛的实质性共识。在本例中，评分集中在 4–5 分的比例较高，同时完全匹配率也较高（标注者对之间为 35.7%–78.6%），且在 28.6%（ALE）至 36.8%（ALD）的查询上达成了三方一致，这表明标注者认同这些查询有意义，即使他们在通过 5 分量表表达认可时的热情程度略有不同。

实用性。 专家还对 SPARQL 结果表作为综合知识成果的实用性进行了评分。所有查询的平均实用性得分为 $4.08/5$，其中 69% 的评分为 ≥ 4，且超过半数为最高分 5。逐查询来看，33 个查询中有 25 个的平均实用性得分至少为 4，只有两个查询的平均实用性低于 3。ALD 查询的实用性得分略高于 ALE 查询（平均值分别为 4.16 和 3.98）。图8（面板 8a 和 8b ）显示，实用性评分的分布比有意义性评分略为分散，尤其是少数表格稀疏或缺少关键参数的查询；不过，大多数数据点仍落在 4–5 分区间。

在 ALD 查询中，12 个查询的平均有用性评分达到满分 5.00（Q01、Q02、Q04、Q05、Q06、Q13、Q14、Q15、Q16、Q17、Q18、Q19），而评分最低的是 Q08，平均分为 3.00。对于 ALE，评分最高的是 Q06，平均分为 4.67；评分最低的是 Q12（平均分 2.67）和 Q14（平均分 3.00）——这是两个领域中仅有的平均分低于 3.00 的查询。从单个标注者的层面来看，ALD 的平均有用性评分介于 4.21 至 5.00 之间，ALE 则介于 2.93 至 4.36 之间，这表明评估者在有用性标准上的严格程度差异，比在有意义性标准上的差异更大。

正式的标注者间一致性再次较低：对于 ALD，Fleiss' $\kappa = -0.035$ 和 $\alpha_{\text{ordinal}} = 0.018$；对于 ALE，$\kappa = 0.151$ 和 $\alpha_{\text{ordinal}} = -0.041$。在 ALD 的案例中，这些数值主要受到天花板效应的影响：三位标注者中有两位几乎给所有查询都打了最高分，因此统计指标可捕捉的变异很少，尽管 63.2% 的查询体现出三方完全一致。ALE 的有用性评分分散程度更高，其中一位标注者给出的评分系统性地低于另外两位（平均分分别为 2.93、4.36 和 4.21）。因此，三方完全一致率降至 21.4%。两位评分较高的标注者之间仍有较高的成对一致性（完全匹配率为 78.6%，Cohen's $\kappa = 0.576$），这表明总体一致性偏低，主要是由于对何为"有用"的综合存在不同的基线预期，而非对标准的应用不一致。

综合来看，这些定量结果表明，领域专家不仅认为这些问题有意义，也确实认为 ORKG 比较生成的机器可操作的综合结果有助于他们推理工艺、机制以及材料–性能权衡。

定性反馈。 对于较低实用性评分的自由文本说明，主要指出当前数据建模和覆盖范围的局限，而非机器可操作的综述这一概念本身存在问题。专家经常要求补充参数（例如黏附系数、表面取向和终止方式、更明确的前驱体／共反应物区分、明确列出掺杂剂的列、每周期生长量的误差棒或标准差，以及每个 ALD/ALE周期更详细的时间参数），更一致地对机制和前驱体系列进行分类，并改进结果呈现方式（例如删除重复行、压缩过长的表格，或解释派生指标，例如稀土MOSLED示例中的效率指标）。在若干情况下，低评分归因于 ORKG 比较内容不完整或填充稀疏，或归因于某些查询——尽管在技术上正确——仍过于笼统，无法支持具体决策。这一模式表明，不满主要集中在具体的数据建模选择和缺失的注释上，而非查询符号化综述知识这一理念。

开放式问题印证了这一解读，同时凸显出热情与务实方面的担忧。在被要求评价机器可操作的科学知识的总体实用性时，三位专家给出的评分均在 4 到 5 之间（平均值为 4.67），他们强调了更快获取综合见解的价值，以及能够即时比较不同论文中的条件，而不必像以往那样从多篇综述中手动重建电子表格。他们提出的"愿望清单"式问题进一步延伸了这一主题：例如，查询特定材料最常见的表面取向和表面终止；按照前驱体/共反应物、温度、等离子体功率和反应器类型，对 ALD工艺配方进行多参数汇总；或者系统综述特定器件架构的工艺窗口和性能指标。换言之，专家们立即将 ORKG 比较投射到更丰富的多维查询上，这些查询反映了他们已有的工艺和材料思维方式。

与此同时，专家预见到更广泛采用仍面临若干障碍。这些障碍包括：（i）除了传统 PDF文件之外，还需要额外投入时间和精力来录入机器可操作的表格；（ii）文献中的报告规范参差不齐，限制了可被一致建模的内容；（iii）如果表格需要（部分）自动提取，人们会担忧数据质量和可信度；以及（iv）缺乏既有的

ALD/ALE 社区标准，尚未明确应记录哪些工艺参数和性能指标。积极的一面是，他们明确指出了一些能够促使其使用或发布此类表格的激励因素：在文献综述中可切实节省时间、实现社区范围内的模式标准化、拥有便于编写与查询 ORKG 比较的用户友好型工具链，以及对其自身工作流程的明确益处（例如能够“一键”回答复杂的比较性问题）。