PRIMAD-LID：一种用于计算可复现性的开发框架

Meznah Aloqalaa[1], Stian Soiland-Reyes[1] & Carole Goble[1]
0000-0002-1382-3896 0000-0001-9842-9718 0000-0003-1219-2137

摘要

过去十年来，随着人们日益关注计算可复现性，界定可复现性的工作也取得了重大进展。然而，这些定义提供的是文字描述，而非框架。学界一直在寻求概念框架，以识别为确保计算可复现性可信而必须加以控制和描述的所有 因素。**PRIMAD**模型最初旨在解决计算可复现性相关术语使用不一致的问题，并概述了六个关键因素：**P**（平台）、**R**（研究目标）、**I**（实现）、**M** （方法）、**A**（参与者）和**D**（数据）。此后，不同领域的多项研究采用了该模型，并提出了扩展版本。然而，这些成果仍然彼此零散，需要进行系统集成和跨学科验证。为弥合这一差距，并认识到 PRIMAD 为计算科学中的可复现性提供了广泛适用的框架，本研究对 PRIMAD 模型展开了深入考察，将其既有扩展整合为一个适用于不同研究情境的统一框架。由此形成了**PRIMAD-LID**，一个用于学科诊断的可复现性框架。该框架保留了 PRIMAD 原有的六个维度，并为每个维度增添三个总体修饰因素：**L**ifespan（时间限定符）、**I**nterpretation（情境推理）和**D**epth（必要粒度），从而为计算可复现性实践奠定了更连贯、更稳健的基础。

[1] 该　曼彻斯特大学计算机科学系，曼彻斯特，M13 9PL，英国

引言

可复现性是科学研究的一个基本方面，它确保研究发现的有效性，并使其他研究人员能够有信心地复现和拓展这些发现。“可重复研究”一词最早于20世纪90年代初出现，由地球物理学家 Jon Claerbout 提出 (Claerbout &Karrenbach, 1992)。此后，文献中出现了许多定义 (Donoho 等，2009；Peng，2011； Stodden，2014 年； Goodman 等 ， 2016 年）。NASEM（美国国家科学、工程与医学科学院）发布的一份重要共识报告（ 美国国家科学院 ， 2019年）将计算可复现性定义为“使用相同的输入数据、计算步骤、方法、代码和分析条件，获得一致计算结果的能力”；将可重复性定义为“在旨在回答同一科学问题的不同研究中获得一致结果的能力，而每项研究都使用了各自的数据”。

NASEM 的定义在讨论可复现性时，隐含地援引了固定性原则。实质上，这些定义承认，实验的某些方面应保持固定，而在比较结果时，其他方面则可以有所变化。此外，这些定义明确指出了应当能够实现的结果；但它们主要仍是文本描述，本身并未规定在实践中实现可复现性的操作方案。具体而言，可复现性往往取决于发布并保存一系列制品和决策，例如：（1）带版本控制的源代码和带标签的发布版本（如 Git 提交哈希和发布存档）；（2）用于捕获执行环境的容器或虚拟机映像（如 Docker/Singularity 配方）；（3）每次运行所用的精确参数和配置文件（如用于模型训练和推理的 YAML/JSON 配置）；以及（4）数据溯源和版本控制（数据集标识符、校验和及预处理步骤）。

PRIMAD 框架通过系统地从六个核心组件——平台、研究目标、实现、方法、参与者和数据——对计算实验进行描述，明确地将同一种关于“固定性”的基本概念转化为可操作的形式（Freire 等人，2016 年）。换言之，NASEM 提供了以结果为重点的可复现性和可重复性的文本定义，而 PRIMAD 则提供了基于组件的框架，明确指出在实践中要证实这些主张，究竟必须保持哪些因素不变，或允许哪些因素变化。

原始 PRIMAD 模型

该模型最初是在 2016[2], Dagstuhl 研讨会 16321 期间提出的，旨在帮助研究人员理解他们所评估的可复现性相关术语。在 PRIMAD 框架中，人们会明确指定在某次复现尝试中哪些组件保持不变、哪些组件发生变化，从而清楚说明复现结果的条件。这种结构化方法直接解决了“R 词汇”使用不一致所造成的混淆（De Roure，2014 年）。PRIMAD 模型将每个术语对应到实验中一组特定的固定组件与可变组件，从而消除了术语含义上的歧义。

缩写“PRIMAD”（表 1）代表在复现计算研究过程中可能发生变化（或保持不变）的关键变量（Freire 等，2016）：

- **P：平台：**执行环境和计算环境，涵盖硬件、软件、编译器、存储和／或用途不同的平台（例如用于管理和传播的平台）。

- **R：研究目标**或研究的目标。
- **I：实现：**计算代码或源代码。
- **M：方法：**实现研究目标的算法、伪代码和／或方法学。
- **A：参与者：**为研究作出贡献的人员。
- **D：数据：**源代码的输入数据和参数值。

在 PRIMAD 框架中，计算可复现性研究采用不同的标签。表 1 清晰地将计算可重复性标签与预期收益及实现这些收益所需的维度联系起来。所列维度明确说明哪些要素会发生变化、相应调整或保持不变。(Aloqalaa 等，2024) 对原始 PRIMAD 表格（Freire 等，2016）中的可复现性标签进行了调整和扩展，以便更精确地指导可重复性实践及其目的和结果，并将其称为“*PRIMAD 分类体系*”。例如，“重新参数化”或“重新校准”，需要更改数据参数，以评估系统敏感性或优化。“迁移”则需要更改执行环境，以评估实验可移植性。

[2] https://www.dagstuhl.de/16321

(Freire 等，2016) 在 PRIMAD 模型中引入了非直接维度：**一致性** 和**透明性**，因为它们并未在模型缩写中明确提及。评估可复现性研究成功与否，依据的是结果是否与先前结果一致，而非完全相同。另一方面，透明性体现了检查所有必要组件的能力，以理解从假设到结果的过程。

计算 可复现性 标签	P	R	I	M	A	D		可复现性增益
						Data		
	Platform	Reference Object	Implementation	Method	Actor	Data Warehouse	Data Parameter	
重复	-	-	-	-	-	-	-	跨试验一致性、确定性
Port, 迁移	X	-	-	-	-	-	-	跨平台兼容性（可移植性）， 最小依赖（灵活性）
重复使用/重新利用	-	X	-	-	-	-	-	跨学科应用（应用代码 于不同场景），资源效率
重新编码、重新解释	(X)	-	X	-	-	-	-	提升代码质量、实现 正确性、扩大采用， 提升效率、灵活性
批准，验证	(X)	-	(X)	X	-	(X)	(X)	假设正确性，通过.....进行验证 不同方法，研究发现稳健性
审查，独立验证	-	-	-	-	X	-	-	增强透明度（充分 信息），独立验证
重新抽样，泛化	-	-	-	-	-	X	(X)	适用于不同情境
重新参数化，重新校准，参数扫描，	-	-	-	-	-	-	X	鲁棒性，对其他条件的适应 （敏感性），参数优化

表 1。 PRIMAD 分类体系。计算实验中用于可复现性的 PRIMAD 模型；每个可复现性标签都与其增益及实现方式相关联。哪些维度将发生变化 X、相应地变化 (X)，或不会变化 -（改编自 (Freire 等 人， 2016 年；Aloqalaa 等 人， 2024年）。

-

先前 PRIMAD 模型采用

为了更全面地了解 PRIMAD 模型如何在不同领域中得到采用和扩展，我们对相关文献进行了按时间顺序的审查。表2呈现了此次审查的结果，按发表年份、应用领域、采用 PRIMAD 的主要目的以及是否对该模型进行了扩展，列出了各项研究。这一综合视图突出了

PRIMAD 作为实践中的可复现性框架所发挥作用的多种方式，并为其持续演进提供了见解。

如表2所示，PRIMAD 模型已应用于多个领域，从信息检索（IR）和音乐信息检索到生物医学工作流程，以及天体物理学中的高性能计算、可视化，甚至人机交互；并作为探讨可复现性的基准框架。研究人员通常借助 PRIMAD 来（1）评估和衡量可复现性，（2）识别障碍或缺口，（3）开发工具和指南，或（4）完善模型本身。例如，Ferro 等人（2016 年）和 Ferro 等人（2019年）使用 PRIMAD 衡量信息检索实验的可复现性，系统地考察重新运行实验时，不同实验组件会如何影响结果。Peter Knees等人（2022年）则将 PRIMAD 应用于以用户为中心的音乐信息检索实验，识别出与用户行为相关的缺失要素，并指出了更广泛的可复现性障碍。Fekete等人（2020 还探讨了可视化社区面临的可复现性挑战，使用 PRIMAD 找出文档和不断演进的软件工具方面的缺口。

在可复现性工具领域，Samuel & König-Ries (2022年) 将 PRIMAD 与语义溯源本体相结合，提出了一种基于溯源的方法，用于自动检查完整性。在其他情况下，研究人员将 PRIMAD 与互补框架相结合——例如，将其与数据完整性信任模型对齐（Schreyer et al., 2025 或 Whole Tale（Brinckman et al., 2018年）计算研究对象概念（Chapp et al., 2020）。另一方面，Staudinger et al.（2024年）通过在机器学习和信息检索情境中进行“压力测试”，评估了 PRIMAD 模型：系统地改变模型维度，最终提出了政策层面和基于基础设施的改进，同时不改变 PRIMAD 框架的核心。

在下一节中，我们将重点讨论直接改进 PRIMAD 或提出新扩展的研究。这些研究为本研究进一步扩展模型、解决以往应用中发现的局限性提供了动机基础。

研究（年份）	学科	子学科	PRIMAD 的用途	提出的扩展
（Ferro 等人，2016 年）	计算机科学	信息检索	-评估可复现性	-
(Gryk 和 Ludäscher，2017)	生命科学	生物信息学	-识别可复现性障碍	-
（Freire 与 Chirigati，2018年）	计算机科学	计算科学（通用）	-扩展/完善 PRIMAD 模型	是，新增了覆盖度和持久性至 PRIMAD。
（Ferro 等人，2019年）	计算机科学	信息检索	-评估可复现性	-
(Chapp等，2019年)	天体物理学	高性能计算	-评估 PRIMAD 模型	区分了部分细化 指出方法与实现的区别 强调了参与者面临的挑战 并敦促关注数据集差异 领域特定的细化。
（Chapp 等，2020年）	天体物理学	高性能计算	-开发可复现性工具/指南	-
（Fekete 等，2020年）	计算机科学	数据可视化	-识别可复现性障碍	-
(Breuer，2020)	计算机科学	信息检索	-评估可复现性	-
(Schaible 等人，2020)	计算机科学	数字图书馆和信息检索平台	-识别可复现性障碍 -比较领域	未作正式扩展，指出 缺少“用户”这一要素。
(Breuer 等人，2022)	计算机科学	信息检索	- 开发可复现性工具/指南	-
(Breuer，2023年)	计算机科学	信息检索	-扩展/完善 PRIMAD 模型 -开发可复现性工具/指南	是，提出了 PRIMAD-U，新增了“用户”作为一个新组件。
（P Knees 等人，2022年）	计算机科学与音乐学	音乐信息检索	-识别可复现性障碍	没有正式扩展指出存在 与用户因素有关的差距。
（塞缪尔 & 柯尼希-里斯，2022年）	计算机科学	计算科学	-开发可复现性工具/指南	-
(Aloqalaa 等，2024)	生命科学	生物信息学	-评估可复现性 -扩展/完善 PRIMAD 模型	Yes, 发现了 PRIMAD 中的遗漏， 建议进行扩展，包括 “维度深度”
（Kern，2024年）	计算机科学	信息检索和机器学习	-识别可复现性障碍	-
(Staudinger 等，2024年)	计算机科学	信息检索和机器学习	-压力测试 PRIMAD	-
(Schreyer 等，2025)	计算机科学	人机交互与数据管理	-开发可复现性工具/指南	-

表2。文献中 PRIMAD 模型的时间顺序概述，以及其在各学科中的应用、目的和扩展。

PRIMAD-LID 的动机与构建

在可重复性文献中，人们普遍认识到，可复现性并非二元概念，而是一个连续谱（Peng，2011；Goodman 等，2016；Tatman 等，2018；Gundersen，2021；Feger 与 Woźniak，2022；Raghupathi 等，2022）。尽管已有这一认识，流传的众多可复现性定义仍常常缺乏一致性，阻碍了可复现的实验结果的实现（FASEB，2016）。多位学者从不同角度应对了这些语义挑战。例如，Plesser（2017）对 1992 至 2017 年间发表的 18 种可复现性定义进行了历史回顾；Lorena（2018）则将 43 种定义归类，区分了可复现性与可重复性。最后，Gundersen（2021）通过对 35 篇学术文献（1992—2020 年）进行系统性调查，深化了对计算可复现性的理解，如图 1 所示。

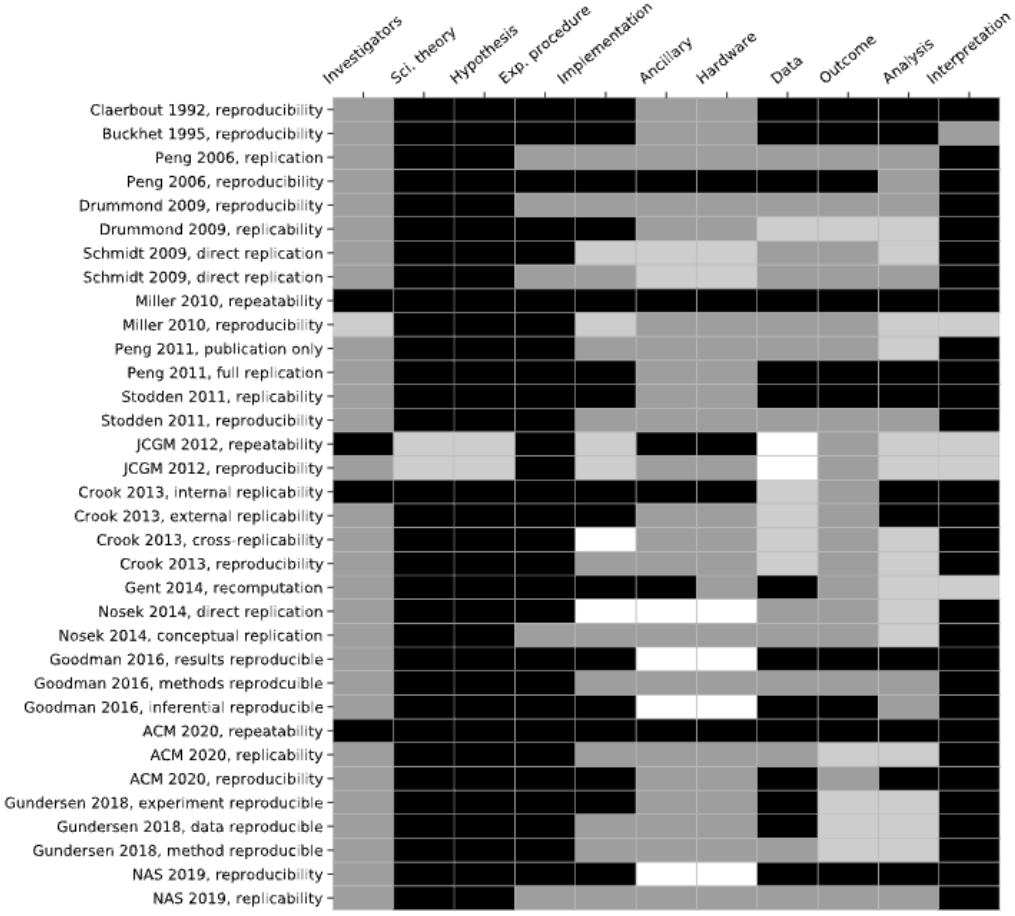

图 1. Gundersen 可重复性定义术语调查：每个单元格的颜色表示原始实验与可重复性实验之间哪些内容必须保持相同（黑色）、相似（白色）、不同（深灰色）或未作说明（浅灰色）。直接采用自（Gundersen，2021）

根据这项调查，Gundersen 确定了影响可复现性的 11 个因素：**研究人员、科学理论**, **假设**, **实验程序**, **实现**, **辅助软件**, **硬件**, **数据**,**结果**, **分析**, and **解释**。每个因素都根据其在所审查的定义中是保持不变、发生变化，还是未作说明来加以描述。随后，Gundersen 将可复现性定义为“独立研究人员遵循原研究人员共享的文档，能够从实验中得出相同结论的能力”。值得注意的是，“结论”由三个术语决定——结果、分析和解释；其中，结果指实验的产出，而解释则源自对结果的分析（Gundersen，2021）。基于这些区分，他提出了可复现性的三个层次：结果可复现性、分析可复现性和解释可复现性。与许多早期定义相比，Gundersen 的界定更为连贯、准确，能够确定复现过程的程度，从而区分结果本身、对结果的分析以及对结果的解释。然而，这一界定仍主要停留在文字描述层面，缺乏结构化框架。

相比之下，PRIMAD 模型提供了一种更全面、以概念为驱动的视角，展示了各种可复现性术语如何映射到计算研究的关键组件上。如前所述，研究人员已在不同领域应用模型维度，以实现特定的计算可复现性目标（见表 1）。此外，一些研究还针对专业应用或更广泛的用途，对模型进行了细化或增加了新的维度，从而进一步增强了该模型。

<table>
<tr><td rowspan="2">Original Model</td><td rowspan="2">PRIMAD Model Dimensions (Freire et al.,2016)</td><td>Implicit*</td><td colspan="8">Transparency</td><td>Consistency</td></tr>
<tr><td>Explicit*</td><td colspan="2">Input | Parameter
Data</td><td>Implementation</td><td>Method</td><td>Platform</td><td>Actor</td><td>Research Objective</td><td></td></tr>
<tr><td>Reproducibility Definition Survey Terms</td><td colspan="2">(Gundersen 2021)</td><td>Data</td><td>Outcome</td><td>Implementation</td><td>Science Theory
Experimental Procedure</td><td>Hardware
Ancillary Software</td><td>Investigators</td><td>Hypothesis</td><td>Analysis | Interpretation</td></tr>
<tr><td rowspan="4">Extensions and Refinements to the PRIMAD Model in the literature</td><td colspan="2">(Chapp et al., 2019)</td><td colspan="2">Dataset Differences</td><td></td><td></td><td></td><td>Actor Role</td><td></td><td></td></tr>
<tr><td colspan="2">(Schaible et al., 2020)
(Knees et al., 2022)
(Breuer 2023)</td><td colspan="2"></td><td></td><td></td><td></td><td>End-User or Participant</td><td></td><td></td></tr>
<tr><td colspan="2">(Freire and Chirigati 2018)</td><td colspan="8">Longevity
Coverage</td><td></td></tr>
<tr><td colspan="2">(Aloqalaa et al.,2024)</td><td colspan="8">Dimension Depth</td><td></td></tr>
</table>

表3. PRIMAD 模型扩展概览，与 Gundersen 的十一项可复现性因素相对应。这一综合视图为本文提出的扩展提供了路线图。

表3总结了这些新增内容，并将其映射到原始 PRIMAD 模型和 Gundersen 的十一项可复现性因素，从而为我们的工作全面呈现了该模型的演变过程。此外，将该模型与 Gunderson 的术语调查相结合，进一步支持了 Freire 等，2016 的建议，即应更深入地考察 PRIMAD 维度及其变化方式，以涵盖不同研究社区所采用的多样化可复现性定义。

除了六个显式维度外，Freire 等人（2016）还提出了 **透明性**和 **一致性**这两个隐式维度。PRIMAD 核心组成部分描述了需要复现的内容，而**透明性**关注这些组成部分的文档记录是否公开且清晰，以确保任何人都能“查看所有必要的组成部分，理解从假设到结果的过程”。这一关注源于这样的认识：即使所有 PRIMAD 组件都已具备，若每个组件都缺少清晰的文档记录，也会阻碍可复现性。另一方面，**一致性，**通过将评估标准从完全相同的结果转变为一致的结果，扩展了 PRIMAD 模型。Gundersen（2021）将可复现性区分为三个层次——结果、分析和解释，它们共同构成计算实验的结论。在 PRIMAD 中，**分析** 和**解释**与一致性相对应，而**结果**则属于数据维度，是实验“数据输出”的一部分，如表3所示。

后来，Freire & Chirigati （2018年） 提出了 **持久性** 和 **覆盖度**作为额外维度，并强调了 **透明性** 的重要性。如表3所示，这些隐式维度嵌入在 PRIMAD 的六个显式维度之中。**持久性**源于这样一种认识：实验必须在最初创建很久之后仍可复现，并应应对软件衰退、弃用的依赖项或数据丢失等挑战。与此同时，**覆盖度**沿袭了 Freire 等 人 （2012年）提出的早期观点，关注原始实验在实际可行的范围内能够复现多少。专有或敏感数据、资源限制以及时间流逝等实际约束，可能导致部分可复现性。即使完全透明，某些实验步骤仍可能无法复现，因此衡量覆盖度有助于兼顾完全可复现性与部分可复现性。

Chapp 等人（2019）研究了 PRIMAD 在大规模科学工作流中的应用，特别关注 LIGO 的引力波数据分析。他们通过 **区分方法与实现** 对该模型进行了部分细化，并指出，即使对代码或

参数进行细微修改，也可能在不改变总体方法学的情况下影响可复现性。他们还探讨了记录 **参与者维度** 相关信息所面临的挑战，尤其是在涉及多个团队或人员不断变动的情况下。此外，他们强调了 **数据集差异**，指出数据类型或质量的变化可能会显著影响可复现性。最后，他们呼吁进行 **领域特定的细化**，以便更好地使 PRIMAD 适应高性能计算和天体物理学研究的独特需求。

在交互式信息系统和用户研究中，**最终用户或参与者**与研究人员一样发挥着至关重要的作用。Schaible 等（2020）是最早强调真实用户参与对可复现信息检索系统的重要性的人之一。他们指出，PRIMAD 的参与者维度不仅应包括研究人员，还应包括提供反馈的参与者和最终用户。同样，P Knees 等（2022）强调了在音乐检索中复现实验以用户为中心的必要性，并指出可复现性必须超越受控的实验室环境，以考虑人类行为本身具有的可变性。Breuer（2023）进一步拓展了这一观点，提出了**“PRIMAD-U”，**在原有六个因素之外增设了专门的“用户”维度。这个附加组件涵盖了最终用户特征和行为的差异，例如，若在不同的用户群体或模拟用户群体中重复信息检索实验，可复现性可能会如何变化。不过，原始 PRIMAD 模型指出，参与者维度涵盖实验中涉及的任何人员。

Aloqalaa 等（2024）提出的最新完善方案是**维度深度**这一概念。他们采用 PRIMAD 和 BioCompute 对象（BCO）标准研究了生物信息学流程的可复现性（Simonyan 等人，2017；Alterovitz 等人，2018）。BCO 是一项 IEEE 2791-2020 标准，适用于由高通量测序（HTS）生成的生物信息学分析。这项研究表明，尽管 PRIMAD 为工作流提供了稳健的高层级结构，但仍需要对每个维度作出更细致的规范。在分析一个基因组测序流程时，研究人员将 PRIMAD 映射到 BioCompute 丰富的元数据时发现了缺漏和歧义，这凸显了深入细化每个 PRIMAD 组件的必要性。维度深度通过十项跨领域属性扩展了 PRIMAD 的每个维度：时间、额外资源、数据分类、功能、版本、人类角色、许可证、容错、审查，以及元数据模式。因此，PRIMAD 从一组因素转变为

一份内容更丰富的清单。每个维度实际上都成为包含详细字段的标题，使模型更加全面，也更具学科针对性。

因此，如表3所示，要将隐含维度（如一致性和透明性（Freire 等，2016）、持久性和覆盖度（Freire 与 Chirigati，2018），以及维度深度（Aloqalaa 等，2024））与用户/参与者组成部分（Schaible 等人，2020；P Knees 等人，2022；Breuer，2023）等额外改进，以及 Chapp 等人（2019）的建议有效整合起来，就需要将这些建议系统地纳入模型的主要维度中。这一过程还包括整合 Gundersen（2021）的可复现性调查术语，从而确保构建一个连贯的框架，涵盖广泛的可复现性因素。

PRIMAD-LID 框架

我们提出“PRIMAD-LID”作为对原始 PRIMAD 模型的增强。这个新缩写保留了 PRIMAD 的六个核心维度：**平台**、**研究目标**、**实现**、**方法**、**参与者**和**数据**，同时为每个维度增添三个贯穿各维度的修饰项：“**L**：生命周期”“**I**：解释”和“**D**：深度”。因此，额外的“**LID**” 修饰项与原有 PRIMAD 维度垂直对齐。这种对齐方式为考察可复现性提供了一个更全面、简洁的框架，后文将对此进行阐述。

L：生命周期维度

确保软件持久性是维持长期计算可复现性的关键因素。除了 Freire 与 Chirigati（2018）提出的、作为 PRIMAD 模型扩展的持久性概念外，美国国家科学院（2019年）还强调，数据与代码存储过时可能会对可复现性构成严重障碍。随着操作系统不断演进，软件库和依赖项发生变化，代码库可能被撤下或迁移，要长期维持软件一致性便成为一项艰巨任务（Kuttel，2021）。这一挑战在依赖相互关联的工具和数据处理流水线的研究工作流中尤为突出。

在这一扩展中，我们将**生命周期维度（L）**定义为实验要素的时间限定符，尤其是 PRIMAD 中的方法、平台、实现和数据。各种时间指标，例如发布日期、更新日期、版本历史和禁运期，揭示了每件制品在多长时间内仍然有效且可正常运行。通过追踪这些信息，研究人员可以判断，例如，

脚本或数据集在几年后是否仍然可用（Grayson、Milewicz 等，2023年）。

生命周期维度可以从历时年龄和功能年龄两个方面来看。历时年龄衡量制品自创建以来经过的时间，而功能年龄则评估其持续的实用性和兼容性。历时年龄较高并不一定意味着已经过时。例如，天文学界至今仍在使用于1981年推出的灵活图像传输系统（FITS）格式（Borgman & Wofford，2021）。同样，GitHub 作为广泛采用的代码仓库，已服务超过15年；而 Google Code 等其他服务则已停止运营，尽管它们的生命周期相近甚至更短。另一方面， 功能年龄取决于服务成熟度、社区采用情况以及为确保技术兼容性而持续更新等因素。因此，研究人员选择平台和格式时，必须着眼于可持续性。在关于研究软件注册库和代码库的九项最佳实践中，Garijo 等人（2022年）建议制定保留和生命周期终止政策，以支持长期可用性。

现实世界的研究进一步说明，关注生命周期如何提升可复现性。例如，(Grayson, Marinov, et al., 2023) 考察了工作流管理平台 Nextflow (P. A. Ewels et al., 2020) 和 Snakemake (Köster & Rahmann, 2012)，研究范围涵盖最多五年的修订。他们发现，主动精心整理更新可使工作流的“完整性”随时间提高三倍。在自然语言处理领域，代码库迁移、失效链接以及不断演变的数据标准，都会随着时间推移降低数据集和软件的可访问性 (Mieskes, 2017)。因此，需要持续减轻时间流逝带来的影响，以促进实验整合和长期可复现性。研究人员强调了持续进行环境管理以应对这一问题的重要性：(Ziemann et al., 2023) 将对计算实验执行环境的持续控制称为“计算环境控制”，而 (P. Ewels et al., 2019) 则用计算环境的持续集成来描述类似实践。这两个概念都强调主动维护平台和软件组件，确保随时间发生的变化不会损害可复现性。

由于实验组件的生命周期无法简化为一个单一数值，我们建议报告一系列日期：发布日期、修改日期、最后访问日期和审查日期，此外还应报告

版本历史日志，以反映其持续可用性。例如，BCO 元数据会记录整个流水线的过时日期，以及每项资源的禁运、创建、修改、访问，和审查日期（Simonyan 等人，2017；Alterovitz 等人，2018）。然而，仅记录制品日期或版本号并不足以确保实验的持久性。其他关键实践同样不可或缺，例如建立开源环境、以标准化格式系统记录数据、代码和结果，以及将叙述性说明与分析关联起来（**诠释维度**）。Akhlaghi 等人（2021）提出了计算实验持久性的八项标准：完整性、模块化、最低复杂度、可扩展性、记录的历史、与分析关联的叙述，以及自由/开源软件（*FOSS*）。这些标准表明，可复现性并非只是重建最初进行的实验，而是要辨别哪些维度属性保持稳定，哪些则必须针对实验的新背景或目的进行更新或替换。在 PRIMAD-LID 框架中，生命周期维度体现了这一理念：每个组件的持久性都可能单独或整体影响可复现性。

I：诠释维度

学术研究日益认识到，可复现性还需要一个额外的维度：对结果的诠释。诠释层面关注的是，独立研究者能否根据一项研究的发现得出相同的结论和解释（Peng，2011；Goodman 等，2016；美国国家科学院，2019；Tatman 等，2018；Gundersen，2021）。PRIMAD 模型旨在对复现过程中可能发生变化的实验组件进行分类。然而，拓展 PRIMAD 的专家强调，可复现性“本身绝非目的”（Freire 等人，2016）；相反，其旨在产出能够支持科学主张的知识或证据。例如，在保持研究问题不变的情况下更改方法，可以“使用不同的方法途径验证假设的正确性”，从而对结论进行独立于方法的检验。在*PRIMAD* 分类体系中，这种检验称为“批准”或“验证”（表 1）。本质上，这是对诠释进行复现：如果不同的方法得出一致的结果，就能增强人们对原始结论的信心。

此外，原始 PRIMAD 框架将**一致性**视为隐含维度，并建议根据可复现性研究的结果与原始发现的契合程度来判断其是否成功，而非要求逐位完全一致。在此基础上，引入一个

诠释维度（I）并将其纳入 PRIMAD-LID 是顺理成章的一步。该维度涵盖一致性，并采用 Gundersen（2021）关于分析和解释的术语来阐明结论的得出过程，如表3所示。重要的是，诠释也兼顾了可复现性中领域特定的情境化与论证。它确保对结果的解释、科学推理和假设验证等因素都得到明确考量。将**诠释维度（I）**作为修饰维度纳入 PRIMAD 各维度后，研究人员会被促使不仅记录并共享数据和代码，还要记录并共享实验要素背后的推理、决策与选择。此外，这也鼓励复现研究者报告他们是否采用了相同的流程、得出了相同的结论，或发现了解释上的差异。

D：深度维度

记录实验组成部分所需的详细程度，即“深度”，通常因领域而异，因为不同研究社区对何为可复现的结果各有自己的标准。FAIR原则强调了这一点，指出元数据必须包含详细的溯源信息，并符合相关领域社区的标准（Wilkinson 等人，2016 年）。因此，可复现性工作不能采用“一刀切”的方法（Stewart 等人，2021）。

除了原始的 PRIMAD 框架（Freire 等人，2016 年）外，Freire 与 Chirigati（2018 年）还强调**透明性** 是一种隐含维度，贯穿所有模型组件。早在 2012 年，Freire 等人便引入了“深度”这一概念，用以说明应记录或共享多少有关计算实验的信息。如前所述，研究者通过列出 10 个详细方面，在生物信息学领域探讨了“深度”（Aloqalaa 等，2024年）。一些领域特定的细节已纳入其他扩展后的 PRIMAD 维度：生命周期中的时间和版本详情，数据中的数据分类，以及诠释维度下的容错（变异接受度）。此外，资源详情、功能、人类角色、许可和元数据等方面，也有助于更全面地呈现计算研究的复杂性。

通过添加**深度维度 （D）**，研究人员可以根据所在领域的标准调整其元数据的粒度，无论是采用生物信息学溯源框架，还是机器学习可复现性检查清单。这种与社区规范的一致性提高了成功

复现的可能性。从本质上说，系统地明确每个 PRIMAD 组件所需的信息深度，可以使模型更加稳健且适应性更强，并确保为各种可复现计算实验提供一个全面的框架。

PRIMAD-LID覆盖度

覆盖度最初由（Freire 等人，2012年）作为一项关于实验可复现性的特征提出。后来，人们提出了一个隐含的"PRIMAD **覆盖度**"维度，用以表示原始实验结果中可复现的部分（Freire 与 Chirigati，2018年）。从本质上说，覆盖度对应于实验产出或结果中可复现的部分。因此，覆盖度是可复现性本身的一种特征（即应用该模型所得到的结果），而非实验设置的描述性方面。与描述实验如何开展或报告的核心 PRIMAD 维度（例如平台、实现、方法、数据等方面的细节）不同，覆盖度表明了可复现性的成功程度，说明实验有多少部分已经或能够被复现。

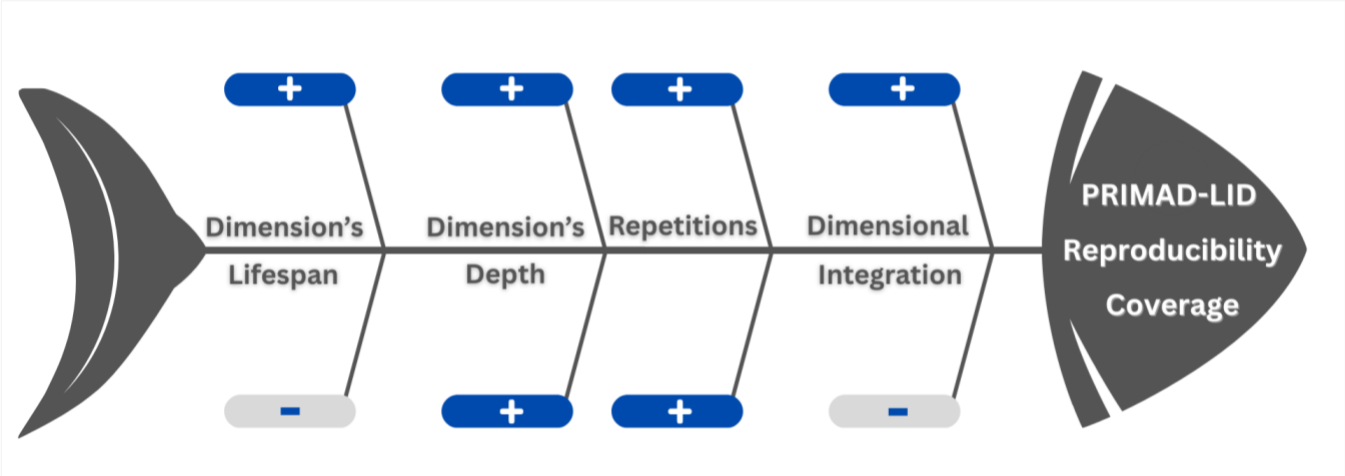

图*2. PRIMAD-LID* 框架可复现性覆盖度的影响图。该图说明，框架维度的深度以及该研究此前的重复次数会对覆盖度产生正向影响。相比之下，维度之间的集成及其生命周期会根据具体情境产生双向影响（正向或负向）。

PRIMAD-LID覆盖度除了受原有PRIMAD维度影响外，还受另外三个因素影响：**深度维度 (D)**、**生命周期维度 (L)**和**维度整合 (T)**。此外，实验的**重复次数 (R)**（重复的频率和范围）也

会显著影响覆盖度，见图2。这些因素都会影响实验能够被多大程度上完整复现：

深度维度（D）：模型维度中更高的深度能够减少歧义空间，从而提升可复现性覆盖度。例如，记录详尽的方法学（对算法、参数和规程进行充分描述）或注释丰富的数据集，能够为未来的研究人员提供更清晰的指导，进而提高实验得以复现的可能性。

生命周期维度（L）：生命周期表示实验的某个组件能够保持可用、可使用或处于最新状态的时长。生命周期对覆盖度的影响较为复杂，并非始终是积极或消极的；其影响可能取决于所考察的维度（例如 GitHub）。较长的持久性能够提升可复现性覆盖度，因为该平台（及其托管的材料）很可能在较长时间内保持可访问和正常运行，使他人即使多年后也能复现实验。相比之下，代码组件"实现"具有较长的生命周期（例如多年保持不变的脚本或软件），可能表明其停滞、存在过时风险或可复用性降低。实证证据表明，在此需要保持谨慎：一项关于软件演化的研究发现，单行代码在被修改或删除前的生命周期中位数约为 2.4 年（Spinellis 等人，2021）。这一相对较短的生命周期中位数意味着，代码在达到某个阶段后很可能会发生变化；如果没有变化，则可能反映出设计和依赖项已经老化。因此，更长的生命周期对于覆盖度而言并非天然是好事或坏事；其影响取决于具体情境。稳定且维护良好的资源能够提升可复现性，而老化且无人维护的资源则可能阻碍可复现性。

维度整合 (T)： PRIMAD 模型的基础在于实验各个维度之间的相互关联，其中某一维度的规范或变更可能会显著影响其他维度（见表 1）。某些维度之间往往存在紧密联系，尤其是方法、实现和平台之间（Aloqalaa 等，2024；Chapp 等，2019）。整合指的是不同实验维度之间耦合或依赖关系的程度。整合度高意味着各个组件（数据、代码、方法、平台等）紧密关联，通常会被打包在统一环境中或在其中执行。整合度低（或分布程度较高）则意味着这些

组件彼此更加独立，可以分开。紧密整合的实验（T 值高）会将代码、数据和环境打包在一个自成一体的设置中，因此易于短期复现且覆盖度较高，但长期来看较为脆弱，因为一旦平台或关键依赖关系出现问题，整个实验都会面临风险。松散整合的实验（T 值低）会将这些组件拆分，使数据、代码和环境能够随时间独立演进、复用和调整，这有利于长期灵活性，但会让日后完整重建原始实验变得更加困难（Akhlaghi 等人，2021）。因此，整合程度会改变两者之间的平衡：紧密整合有助于即时复现，松散整合有助于长期适应性，而整体 PRIMAD-LID 覆盖度也会相应变化。

重复次数（R）与 PRIMAD-LID 覆盖度之间存在明显的**正向（+）**关系。它指实验重复的次数，以及重复实验所处的独立条件数量。每一次成功的重复，尤其是由独立研究者或在不同情境下完成的重复，都会拓展结果成立的条件范围，从而提高再次成功重复的可能性。卡尔·波普尔爵士曾著名地指出，科学发现只有经过多次检验才值得信赖：“即使是我们自己的观察……在重复并检验之前，我们也不会完全认真对待。只有通过这样的重复，我们才能确信自己面对的并非单纯孤立的‘巧合’。”（波普尔，2005）

总之，PRIMAD-LID 模型中的**覆盖度**体现了原始实验在多大程度上能够被复现。它并非绝对属性，而是受到以下因素影响：每个组件的描述深度、每个组件保持有效或可用的时长、组件之间的耦合方式，以及实验重复的次数。要最大限度地提升 PRIMAD-LID 覆盖度，必须谨慎平衡深度、生命周期 和集成。这需要提供丰富的细节、维持可持续的资源，并选择最佳的集成程度，以确保实验不仅现在可复现 ，未来也能保持稳健 。此外，重复实验能够提高覆盖度，因为这会增加每一次新复现成功的可能性，从而增强实验随时间推移的可复现性。

结论

FASEB 圆桌讨论的参与者指出，缺乏统一定义是阻碍实验结果可复现性的一项主要因素（FASEB，2016 年）。与此相反，英国

可复现性网络（UKRN）强调，没有任何一种单一、通用的方法能够完全解决不同情境下的可复现性问题（Stewart 等，2021）。PRIMAD 模型针对这些相互矛盾的建议，提出了一个支持多种可复现性目标的统一框架（Freire 等，2016）。然而，文献中的应用表明，有必要对其进行更全面的扩展。

我们提出 **PRIMAD-LID**，这是对原始 PRIMAD 模型的扩展，新增了三个维度：**生命周期、解释和深度。**这些维度与原有的六个维度相互垂直（**平**台、**研**究目标、**实**现、**方**法、**数**据和**行**动者），有助于厘清并统一该模型在各计算科学领域中的可复现性定义。通过系统整合生命周期、解释和深度，扩展后的模型消除了原框架经完善后可能产生的歧义，确保每个新增维度都有明确的定义，并且能够在实践中落实。该模型突显了确保可复现性所需的关键组件，并可根据不同学科进行调整。此外，PRIMAD-LID 提供了计算实验的模块化表示，这对确保实验的持久性至关重要（Akhlaghi 等人，2021）。因此，它可以作为计算可复现性的路线图：指导研究人员撰写报告，协助开发人员组织工作流，为出版商提供清晰的核对清单，并为利益相关方制定研究数据管理指南。不过，PRIMAD-LID 框架仍需在不同计算领域中进一步检验。这是因为各维度的深度会因所采用的具体方法而异，且在不同学科之间以及同一学科内部都可能存在显著差异。

参考文献

Akhlaghi, M., Infante-Sainz, R., Roukema, B., Khellat, M., Valls-Gabaud, D., & Baena Galle, R.（2021）。迈向长期且可归档的可复现性。*Computing in Science and Engineering*，1–1。https://doi.org/10.1109/MCSE.2021.3072860

Aloqalaa, M., Soiland-Reyes, S., & Goble, C.（2024年）。系统性考察可复现性：使用 PRIMAD 模型和 BioCompute 对象开展高通量测序案例研究。*ArXiv*。https://doi.org/10.48550/arxiv.2412.07502

Alterovitz, G., Dean, D., Goble, C., Crusoe, M. R., Soiland-Reyes, S., Bell, A., Hayes, A., Suresh, A., Purkayastha, A., King, C. H., Taylor, D., Johanson, E., Thompson, E. E., Donaldson, E., Morizono, H.,

Tsang, H., Vora, J. K., Goecks, J., Yao, J., ... Mazumder, R.（2018年）。通过标准化传达 HTS 来源信息、分析和结果，推动精准医疗发展。*PLoS Biology*，*16*（12），e3000099。https://doi.org/10.1371/journal.pbio.3000099

Borgman, C. L., & Wofford, M. F.（2021）。从数据流程到数据产品：天文学中的知识基础设施。*Harvard Data Science Review*。https://doi.org/10.1162/99608f92.4e792052

Breuer, T., Keller, J., & Schaer, P.（2022年）。ir_元数据：面向信息检索实验的可扩展元数据模式。国际 *ACM SIGIR* 信息检索研究与发展会议第 *45* 届论文集，3078–3089。https://doi.org/10.1145/3477495.3531738

Breuer, T.（2020）。可复现在线搜索实验。载于 J. M. Jose、E. Yilmaz、J. Magalhães、P. Castells、N. Ferro、M. J. Silva 和 F. Martins（编），信息检索进展：第 *42* 届欧洲信息检索研究会议，*ECIR 2020*，葡萄牙里斯本，*4* 月 *14* 日–*17* 日，*2020* 年，论文集，第二部分（第 12036 卷，第 597–601 页）。Springer International Publishing。https://doi.org/10.1007/978-3-030-45442-5 77_

Breuer, T.（2023年）。可复现信息检索研究：从有原则的面向系统评估到面向用户的实验。*DuEPublico*：杜伊斯堡-埃森在线出版物，德国杜伊斯堡-埃森大学。https://doi.org/10.17185/duepublico/78449

Brinckman, A.、Chard, K.、Gaffney, N.、Hategan, M.、Jones, M. B.、Kowalik, K.、Kulasekaran, S.、Ludäscher, B.、Mecum, B. D.、Nabrzyski, J.、Stodden, V.、Taylor, I. J.、Turk, M. J. 和 Turner, K.（2018年）。面向可复现性的计算环境：捕获"Whole Tale"。*Future Generation Computer Systems*。https://doi.org/10.1016/j.future.2017.12.029

Chapp, D.、Rorabaugh, D.、Brown, D. A.、Deelman, E.、Vahi, K.、Welch, V. 和 Taufer, M.（2019年）。PRIMAD 模型对 LIGO 引力波搜索工作流的适用性研究。计算机系统实用可复现评估国际研讨会第二届会议论文集 - *P-RECS '19*，1–6。https://doi.org/10.1145/3322790.3330591

Chapp, D.、Stodden, V. 和 Taufer, M.（2020）。构建网络基础设施生态系统中的可复现性愿景：利用社区力量。《超级计算前沿与创新》，*7*（1）。https://doi.org/10.14529/jsfi200106

Claerbout, J. F., & Karrenbach, M.（1992）。电子文档为可重复研究赋予了新的意义。*SEG 1992*年技术计划扩展摘要，601–604。https://doi.org/10.1190/1.1822162

De Roure, D.（2014 年）。学术交流的未来。*Insights: The UKSG Journal*，*27*(3)，233– 238。https://doi.org/10.1629/2048-7754.171

Donoho, D. L.、Maleki, A.、Rahman, I. U.、Shahram, M. 和 Stodden, V.（2009）。计算谐波分析中的可重复研究。*Computing in Science & Engineering*，*11*（1），8–18。https://doi.org/10.1109/MCSE.2009.15

Ewels, P.、Peltzer, A.、Fillinger, S.、Alneberg, J.、Patel, H.、Wilm, A.、Garcia, M.、Di Tommaso, P. 和 Nahnsen, S.（2019）。nf-core：社区策划的生物信息学流程。*BioRxiv*。https://doi.org/10.1101/610741

Ewels, P. A.、Peltzer, A.、Fillinger, S.、Patel, H.、Alneberg, J.、Wilm, A.、Garcia, M. U.、Di Tommaso, P. 和 Nahnsen, S.（2020）。面向社区策划生物信息学流程的 nf-core 框架。*NatureBiotechnology*，*38*（3），276–278。https://doi.org/10.1038/s41587-020-0439-x

FASEB.（2016 年）。提升研究可重复性：美国实验生物学学会联合会的建议。美国实验生物学学会联合会。https://www.aai.org/AAISite/media/Public_Affairs/Policy_Issues/NIH_Peer_Review/FASEB_Enhanci ng-Research-Reproducibility.pdf

Feger, S. S., & Woźniak, P. W.（2022年）。可复现性：以研究者为中心的定义。*MultimodalTechnologies and Interaction*，*6*(2), 17。https://doi.org/10.3390/mti6020017

Fekete, J.-D., Freire, J., & Rhyne, T.-M.（2020）。探索可视化中的可复现性。*IEEE ComputerGraphics and Applications*，*40*(5), 108–119。https://doi.org/10.1109/MCG.2020.3006412

Ferro, N., Fuhr, N., Järvelin, K., Kando, N., Lippold, M., & Zobel, J.（2016 年）。提升信息检索领域的可复现性。*ACM SIGIR Forum*，*50*(1)，68–82。https://doi.org/10.1145/2964797.2964808

Ferro, N., Marchesin, S., Purpura, A., & Silvello（2019年）。对神经信息检索模型进行基于 Docker 的可重复性研究。

Freire, J., Bonnet, P., & Shasha, D.（2012年）。计算可复现性：现状、挑战与数据库研究机遇。*2012*年国际数据管理会议论文集 - *SIGMOD '12*，593。https://doi.org/10.1145/2213836.2213908

Freire, J., & Chirigati, F.（2018）。溯源与计算可复现性的不同类型。发表于 Chirigati 主编的《IEEE 数据工程通报》，41(1)，第 15–26 页，IEEE。

Freire, J., Fuhr, N., & Rauber, A.（2016 年）。电子科学中面向数据的实验的可复现性（Dagstuhl 研讨会 16041）。*Schloss Dagstuhl - Leibniz*计算机科学中心有限公司，瓦登/萨尔布吕肯，德国。https://doi.org/10.4230/dagrep.6.1.108

Garijo, D., Ménager, H., Hwang, L., Trisovic, A., Hucka, M., Morrell, T., Allen, A., 软件注册库最佳实践工作组，& SciCodes联盟。（2022年）。研究软件注册库和代码库的九项最佳实践。*PeerJ*：计算机科学，*8*，e1023。https://doi.org/10.7717/peerj-cs.1023

Goodman, S. N., Fanelli, D., & Ioannidis, J. P.（2016 年）。研究可复现性意味着什么？《科学·转化医学》，*8*(341)，1–6。https://doi.org/10.1126/scitranslmed.aaf5027

Grayson, S., Marinov, D., Katz, D. S., & Milewicz, R.（2023年）。Snakemake工作流目录和nf-core注册库中的工作流自动复现。*2023*年 *ACM*可复现性与可重复性会议论文集，74–84。https://doi.org/10.1145/3589806.3600037

Grayson, S., Milewicz, R., Teves, J., Katz, D. S., & Marinov, D. (2023年). 征集：计算实验自动可复现性标准。*ArXiv*。https://doi.org/10.48550/arxiv.2307.11383

Gryk, M. R., & Ludäscher, B. (2017). 工作流与溯源：面向自然科学的信息科学解决方案。《图书馆趋势》，*65*(4), 555–562。https://doi.org/10.1353/lib.2017.0018

Gundersen, O. E. (2021). 可复现性的基本原则。《哲学学报*A*辑》，数学、物理与工程科学，*379*(2197), 20200210。https://doi.org/10.1098/rsta.2020.0210

Kern, B. M. J. (2024年). 改进信息学本科生辍学预测。*2024*年计算机科学教育创新与技术会议论文集，第*2*卷，830–831。https://doi.org/10.1145/3649405.3659472

Knees, Peter, Ferwerda, B., Rauber, A., Strumbelj, S., Resch, A., Tomandl, L., Bauer, V., Tang, F. Y., Bobinac, J., Ceranic, A., & Dizdar, R. (2022年). 关于以用户为中心的音乐信息检索研究的可复现性研究及其重要性。*Zenodo*. https://doi.org/10.5281/zenodo.7316775

Köster, J., & Rahmann, S. (2012年). Snakemake——一种可扩展的生物信息学工作流引擎。生物信息学,*28*(19), 2520–2522. https://doi.org/10.1093/bioinformatics/bts480

Kuttel, M. M.（2021）。永不过期。《自然·天文学》。https://doi.org/10.1038/s41550-021-01402-3

Lorena, A. B.（2018年）。可复现研究术语。arXiv 预印本 arXiv:1802.03311。https://doi.org/10.48550/arXiv.1802.03311

Mieskes, M.（2017年4月）。对自然语言处理社区中数据的定量研究。载于首届 *ACL* 自然语言处

美国国家科学院。（2019年）。科学中的可复现性与可重复性。国家科学院出版社（美国）。https://doi.org/10.17226/25303

Peng, R. D.（2011年）。计算科学中的可重复研究。科学，*334*（6060），1226–1227。https://doi.org/10.1126/science.1213847

Plesser, H. E.（2017年）。可复现性与可重复性：一个混淆术语的简史。《神经信息学前沿》，*11*，76。https://doi.org/10.3389/fninf.2017.00076

Popper, K. (2005)。《科学发现的逻辑》。Routledge。https://doi.org/10.4324/9780203994627

Raghupathi, W.、Raghupathi, V. 和 Ren, J.（2022）。计算研究中的可复现性：一项实证研究。*IEEE Access*：实用创新，开放解决方案，*10*，29207–29223。https://doi.org/10.1109/ACCESS.2022.3158675

Samuel, S. 和 König-Ries, B.（2022）。采用语义方法实现科学实验可理解性和可复现性的端到端溯源表示。《生物医学语义学杂志》，*13*（1），1。https://doi.org/10.1186/s13326-021-00253-1

Schaible, J.、Breuer, T.、Tavakolpoursaleh, N.、Müller, B.、Wolff, B. 和 Schaer, P.（2020）。学术共享任务的评估基础设施。《数据库领域》，*20*（1），29–36。https://doi.org/10.1007/s13222-020-00335-x

Schreyer, V., Bornschlegl, M. X., & Hemmje, M. (2025). 推进人际对话视频记录应用的标注、可视化与可复现归档。《信息》, 16(5), 349. https://doi.org/10.3390/info16050349

Simonyan, V., Goecks, J., & Mazumder, R. (2017). 生物计算对象——迈向生物医学科学计算的评估与验证。《PDA药学科学与技术杂志》 / PDA, 71(2), 136–146. https://doi.org/10.5731/pdajpst.2016.006734

Spinellis, D., Louridas, P., & Kechagia, M. (2021). 软件演化：细粒度元素的生命周期。*PeerJ*: 计算机科学, *7*, e372. https://doi.org/10.7717/peerj-cs.372

Staudinger, M., Kern, B. M. J., Miksa, T., Arnhold, L., Knees, P., Rauber, A., & Hanbury, A. (2024). 任务可复现性：对机器学习与信息检索研究中可复现性问题的调查。*2024*年*IEEE*第*20*届电子科学国际会议（电子科学）, 1–9. https://doi.org/10.1109/e-Science62913.2024.10678657

Stewart, A. J., Farran, E. K., Grange, J. A., Macleod, M., Munafò, M., Newton, P., Shanks, D. R., & UKRN机构负责人. (2021). 改进研究质量：英国可复现性网络机构负责人对研究改进的看法。*BMC*研究札记, *14*(1), 458. https://doi.org/10.1186/s13104-021-05883-3

Stodden, V. (2014 年). 哪个科学理念已经到了该退休的时候？ Edge. https://www.edge.org/response-detail/25340

Tatman, R., VanderPlas, J., & Dane, S. (2018). 机器学习研究可复现性的实用分类体系。

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016 年). 科学数据管理与治理的 FAIR指导原则。*Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Ziemann, M., Poulain, P., & Bora, A. (2023年). 计算可复现性的五大支柱：生物信息学及其他领域。生物信息学简报，*24*(6)。https://doi.org/10.1093/bib/bbad375