SUPPLEMENTARY MATERIAL

7 RESEARCH METHODS

7.1 PRISMA Literature Review

We adopted two strategies Salvador Oliván et al. (2023) to conduct our survey. On the one hand, active search for articles highly cited reports, and follow-up of citation threads on computational reproducibility. On the other hand, we applied the PRISMA 2020 Haddaway et al. (2022) methodology with the SCOPUS & WoS databases between 2020 to the present day (Table 6). Eventually, we reduced our analysis to 100 representative works. Figure 9 shows the overall flow.

		PRISMA
Step	Num items	Condition/ Query
Identification	SCOPUS (413)	TITLE (reproducibility) AND PUBYEAR > 2019 AND PUBYEAR < 2024 AND (LIMIT-TO (SUBJAREA , "COMP"))
Identification	WoS (371)	TI=(reproducibility) and 2023 or 2022 or 2021 or 2020 (Publication Years) and Multidisciplinary Sciences or Computer Science Interdisciplinary Applications or Computer Science Theory Methods or Computer Science Information Systems (Web of Science Categories)
Deleting repeated	(144)	Repeated articles
Screening	(80)	Main theme reproducibility
Included	(60)	Systematically classified

Table 6. The PRISMA screening was used in this review to select recent relevant articles related to reproducibility in scientific research.

7.2 Journals Survey

To avoid bias in the research and not only address journals that are known to apply reproducibility policies, e.g. IEEE with CodeOcean [50], ACM with reproducibility Badges [51], Journal Nature[52], a request for participation was sent to several journals specialized in computer science.

Information was voluntarily requested through a 16-question form from a list of 500 journals specialized in computer science and Scopus indexed from Q1 to Q4 in the period from July 1, 2023, to September 1, 2023. Although some of the questions were very precise, they were left open to comments.

[50] https://innovate.ieee.org/ieee-code-ocean/

[51] https://www.acm.org/publications/policies/artifact-review-badging

[52] https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards

8 APPENDIX

Tool	Neptune	Weights & Biases	Comet	Sacred & Omni board	MLflow	Tensor Board	Guild AI	Polyaxon	ClearML	Valohai	Pachyderm	Kubeflow	Verta.ai	SageMaker	DVC
Focus	MedaStor, Track, ModReg	Exp Manage	Exp Manage	Exp Manage	Entire Life cycle	Exp Manage	Exp Manage	Exp Manage	Exp Manage	Entire Life cycle	Entire Life cycle	Run Orchestration	Entire Life cycle	Entire Life cycle	Data Versioning
Price	Ind: Free, Acad: Free, Team: Paid	Ind: Free, Acad: Free, Team: Paid	Ind: Free, Acad: Free, Team: Paid	Free	Free	Free	Free	Paid	Free or Paid	Paid	NA	Free	Open-source: Free, Hosted: Paid	Extra cost	DVC: Free, DVC Studio: Free/Paid
Web UI or Console-based?	Web UI	Web UI	Web UI	Web UI	Web UI	Web UI	Both CLI and Web UI	Both Web UI and CLI	Web UI	Both Web UI and CLI	Both Web UI and CLI	Web UI	Web UI	Both Web UI and CLI	Both Web and Console UI
Code Versions				Limited	Limited		Limited		Limited	Limited		Limited	x		x
Metrics and Losses	x	x	x	Limited	x	Limited	Limited	Limited	Limited	Limited		x	x	x	Limited
Images	x	Limited	N/A					Limited	x			N/A			
Audio	x	x	x			x		x	x		N/A				
Video	x	x						x	x		N/A				
One-command Experiment Re-run		x			x		Limited	x	x	x	N/A	N/A			x
Experiment Lineage		x	Limited					Limited		x	x	Limited			
Environment Versioning	x	x	x		x		x	x	x	x	N/A	N/A			x
Saving/ Fetching/ Caching Datasets							Limited		x	x	N/A	N/A			x
Comparing Experiments															
Table Format Diff	x	x	x				x		x	Limited					x
Overlayed Learning Curves	x	x	x	x	x	x		x	x	x				x	x
Code	Limited	Limited		Limited			Limited		x						x

Table 7. Reproducibility Benchmark of MLOps Tools https://neptune.ai/ (2021). This table shows benchmarks for different popular Data/Model lifecycle management tools.

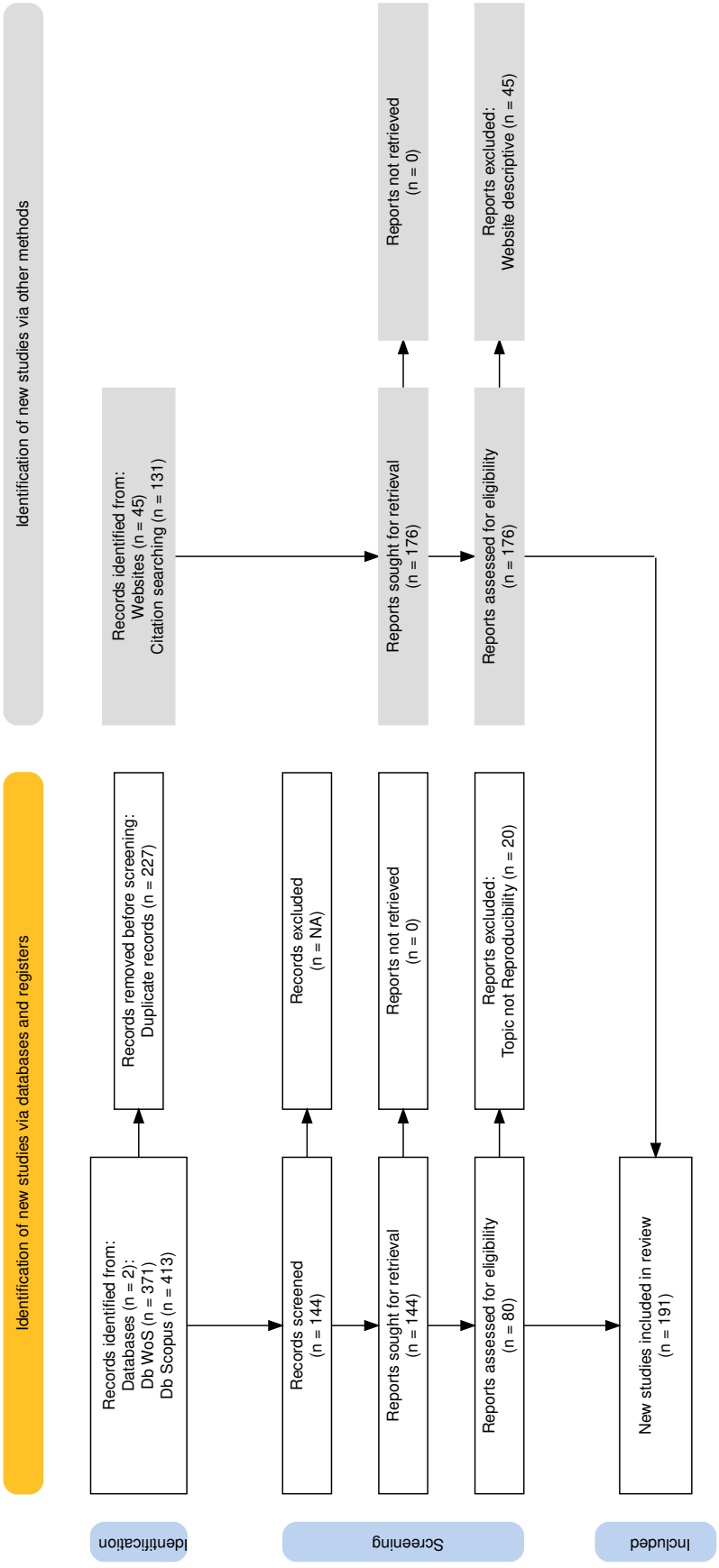

Figure 9. PRISMA flow diagram.