CLAIMED, SHARED, VERIFIED: AUDITING DATA AVAILABILITY IN QUANTUM AND CLASSICAL COMPUTING RESEARCH

Anonymous authors
Paper under double-blind review

ABSTRACT

Data availability is usually summarized as a single headline figure: the share of papers that release their data. Such a figure says how much data is shared, not what holds it there, nor whether stated availability translates into actual access. We bring two complementary approaches to this problem across quantum computing and classical computing research. The first is a large-scale automated classification of arXiv papers published since 2021, scoring each paper's use of external and self-generated data independently across six cases, ranging from directly accessible to claimed-but-unreachable. This lets us compare how the two fields actually share data, rather than how often they say they do. The second is a fine-grained verified-accessibility framework that distinguishes declared availability from independently confirmed access, recording not just whether data exist but whether they are locatable, complete, documented, and persistent, dimensions a binary availability measure cannot capture. Linking each paper to its publication venue, funding sources, and citation record, we test whether availability tracks venue policy, whether funder open-science mandates reach the work they cover, and whether shared data is repaid in citations. Together, the two approaches show that data availability is an outcome of the incentives around a paper rather than a property of the discipline it belongs to, and that declared availability systematically overstates verified accessibility. We release the paper-level classifications and linked metadata so these relationships can be tested further.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

REPRODUCIBILITY STATEMENT

We have taken several steps to support reproducibility of the reported experiments...

REFERENCES

A APPENDIX