Information Research - Vol. 31 No. iConf (2026)

AI for research assessment: opportunities and challenges from India

Tatiana Chakravorti, Chuhao Wu, Sai Koneru, and Sarah Rajtmajer

DOI: https://doi.org/10.47989/ir31iConf64130

Abstract

Introduction. Artificial intelligence (AI) is beginning to reshape the infrastructures of scholarly communication, raising important questions about trust and transparency. This study examines researchers' perspectives on AI-driven scientific replication prediction tools as emerging components of research assessment.

Method. Qualitative, in-depth, semi-structured interviews were conducted with 19 faculty and doctoral scholars in India to explore how researchers perceive and engage with the scientific replication AI tool.

Analysis. Our analysis focused on how information practices, institutional incentives, and sociocultural contexts shape the adoption of AI technologies in scholarly work. Interview transcripts were studied using thematic analysis using a collaborative and iterative coding process.

Results. Participants mention that limited funding, infrastructure, and access to advanced tools or high-quality datasets make replicating studies difficult in India. They recognised the value of AI tools for surfacing reproducibility assessments during literature reviews and study design. They advocated for hybrid human–AI systems that balance the efficiency of automation with the nuanced judgment of experts.

Conclusion. This study situates AI replication tools within broader scholarly infrastructures, highlighting the need for design features that enhance explainability, fairness, and cultural sensitivity. It also represents the challenges faced by Indian researchers when it comes to replication and open science.

Introduction

Concerns about reproducibility and replicability have become increasingly prominent across domains, including psychology, economics, and computer science (Baker, 2016b; Chakravorti et al., 2024; Kapoor & Narayanan, 2023; Willis & Stodden, 2020). Causes such as selective reporting, inadequate documentation of methods, lack of access to data and code, and perverse publication incentives have been widely noted (National Academies of Sciences, Engineering, and Medicine et al., 2019; Shrout & Rodgers, 2018). These concerns are compounded by the rise of predatory publishing and, more recently, by the growing prevalence of generative AI in research workflows, both of which complicate the assessment of scientific integrity (Loken & Gelman, 2017; Maxwell et al., 2015; Schooler, 2014).

Communities in open science and science of science have responded with initiatives to promote openness in data, methods, and infrastructures (Nosek et al., 2015, 2016; Obels et al., 2020; Stodden et al., 2014); strengthen statistical practices (Benjamin et al., 2018); and realign research incentives toward integrity (Nosek et al., 2012, 2018, 2021; Uhlmann et al., 2019). At the same time, advances in scholarly big data and AI have introduced computational approaches to evaluate research credibility at scale. Emerging tools leverage machine learning, information retrieval, and natural language processing to estimate replicability of published scientific findings (Marshall & Wallace, 2019; Pawel & Held, 2020; Yang et al., 2020). They also raise critical questions at the intersection of information science and AI. How might researchers engage with such tools? What forms of explainability and transparency are necessary for trust? And how do local research cultures influence adoption?

Ultimately, the usefulness of AI-supported research assessment tools will be judged not only on technical performance but also on their reception by researchers, institutions, and scholarly infrastructures. To date, however, most studies on reproducibility and research assessment have centered on scholars in the Global North (Cova et al., 2021; Fidler & Wilcox, 2018; Mede et al., 2021; Vilhuber, 2020). Influential surveys, such as the 2016 Nature survey (Baker, 2016a) on reproducibility, have underrepresented voices from the Global South. Yet cultural and institutional contexts shape information practices and norms of openness in important ways, as emphasised in recent calls to decolonise open science (Dutta et al., 2021).

This study seeks to broaden discourse by examining perspectives from India, a research ecosystem underrepresented in conversations around reproducibility and open science. We investigate both the obstacles Indian researchers face in adopting best practices for openness and integrity, and their views on the design and deployment of AI-driven tools for research assessment. In doing so, we situate replication prediction as a motivating scenario to ground discussions of how AI might integrate into scholarly infrastructures.

Our work is scaffolded by the following research questions (RQs):

- **RQ1:** What obstacles do Indian researchers face when engaging in best practices of reproducibility, replicability, and transparency?
- **RQ2:** What institutional or technological developments, such as incentives, repositories, or tools, could help to support research best practices?
- **RQ3:** Do Indian researchers see a place for AI-driven research assessment tools? If so, what are the requirements of such technologies?

To answer these questions, we conducted 19 semi-structured interviews with faculty and doctoral researchers from universities across India. We anchored these interviews around a demonstration of a prototype AI tool for replication prediction, enabling participants to reflect concretely on opportunities, risks, and design needs. Our findings highlight enthusiasm for hybrid human–AI

approaches, emphasise the importance of transparency and explainability, and underscore how sociocultural and institutional contexts shape practices around reproducibility.

Related work

The replication crisis

Recent years have witnessed growing concern about the reproducibility, replicability, and robustness of published research findings across a wide swath of the scientific literature (Baker, 2016a). Large-scale replication projects in psychology (Collaboration, 2015), economics (Camerer et al., 2016), sociology (Camerer et al., 2018), biology (Errington et al., 2014), physics (Feger et al., 2019) and beyond have turned up disappointing results. Various factors have been suggested as contributing to this crisis, including selective reporting (Rowe, 2023), p-hacking (Head et al., 2015), poor theoretical design, and unavailability of code and data (Miyakawa, 2020). Authors have also suggested that root causes of low replication rates include lack of well-aligned incentives and failure of peer review processes (Bajpai et al., 2017; Niksirat et al., 2023; Samuel & König-Ries, 2022). Reproducibility and replicability are increasingly of concern in computer science and AI as well (Collberg & Proebsting, 2016; Dacrema et al., 2021; Hutson, 2018; Pimentel et al., 2019; Raff, 2019). These fields are impacted by similar workflow vulnerabilities and incentives misalignment. While the opacity of deep learning algorithms and sensitivity to parameter tuning magnify these challenges (Gohel et al., 2021; Haibe-Kains et al., 2020). While each field has its own norms and conventions, some general best practices to support research integrity have been proposed (Santana-Perez, Pérez-Hernández, et al., 2015; Stodden & Miguez, 2013). Key amongst these guidelines is ensuring transparency through data sharing (Chard et al., 2015) and ample documentation, e.g., environmental specifications required to replicate computational experiments (Bánáti et al., 2015).

Transparency and open science

Motivation for transparency and openness in research dates back to Merton's scientific norms (Anderson et al., 2010) and the notion of scepticism as central to scientific endeavor. Transparency is operationalised through the full disclosure of data, code, methodology, and other research artifacts (Patil et al., 2016; Sokol, 2019). Increasingly, notions of transparency also include the documentation and disclosure of research chronology, hypothesis generation, and any changes to the planned analytic pipeline made throughout the course of the work, e.g., by preregistration (Nosek et al., 2018). These practices have been heralded as the key to enhancing reproducibility and replicability (Crüwell et al., 2019; McKiernan et al., 2016). Optimistically, the past decade has seen an increase in transparent and open research practices, particularly in the social sciences (Freese et al., 2022; Gilmore et al., 2017; Polanin et al., 2020). These practices, however, are still rare in many fields (Bajpai et al., 2017). For example, (Gunzer et al., 2022) analysed 83 articles on AI neuroimaging models from 2000–2020, finding that only 10.15% had open-source code. Similarly, a recent ACM survey found that none of 93 papers surveyed from 2021–2022 had a corresponding preregistration and only one used a dataset that was made openly available (Haim et al., 2023). In a study of open science norms in clinical psychology, while 98% of 100 papers sampled between 2000 and 2020 had some data available, only one provided an analysis script (López-Nicolás et al., 2022). Recently, a limited number of universities have started rewarding researchers whose work aligns with standards of open science and transparency, e.g., (Schönbrodt & Mellor, 2018). In machine learning and AI, transparency has been central to discussions around ethical engagement with emerging technologies (Ahmad et al., 2020; Lu et al., 2019; Pushkarna et al., 2022). Largely, transparency has been engaged as a precursor to the explainability of algorithms and methodology (Iyer et al., 2018). However, authors have also recommended a broad understanding of research transparency in these fields to include transparency of epistemic outputs, questions about who has access, and full consideration of the social context with which technology interacts (Eyert & Lopez, 2023).

AI for research assessment

In tandem with advances in AI, researchers have explored opportunities for AI tools to evaluate published work with respect to openness, reproducibility, and replicability. In 2019, Altmejd et al. (Altmejd et al., 2019) developed and assessed statistical models to predict the outcomes of replication studies. Their models were trained and tested on replications of published findings in experimental psychology and economics. Similarly, Pawel et al. used data from replication projects to generate probabilistic forecasts of replication outcomes (Pawel & Held, 2020). In 2020, Yang et al. (Yang et al., 2020) introduced a model combining machine intelligence with human judgment to assess the likelihood of replication for a given study. Their model was trained and validated on hundreds of manually replicated studies and tested on independent datasets. These forecasts were subsequently assessed for their discrimination, calibration, and sharpness. Although these models lack explainability. Most recently, authors have proposed AI in the form of an artificial prediction market for research assessment (Rajtmajer et al., 2022) which provides more explainability to the users by using prediction markets. Algorithmic agents (bot traders) buy and sell assets corresponding to notional replication study outcomes. Our interviews include a demonstration of this AI as an exemplar to ground the discussion.

Methodology

We conducted 19 semi-structured interviews (Thakkar et al., 2022; Varanasi et al., 2022) to understand challenges to reproducibility, replicability, and transparency for researchers in India across several fields in social science. IRB approval for human subjects research was obtained prior to participant recruitment.

Participant recruitment

We used snowball sampling to recruit professors and PhD students from private and governmental institutions geographically distributed throughout India. We recruited researchers from both the social sciences and engineering. We posted advertisements on LinkedIn and Twitter. The recruitment advertisement contained information about the study, expected interview length, compensation, inclusion criteria, and a link to the screener survey. This survey helped us to select a diverse sample, with respect to demographics, research domain, and experience. We received a total of 26 responses, from which we selected 19 participants for our study. The ages ranged from 25 to 64 years (median = 41). Of note, most of our participants represent Government institutions; professors and scholars from private institutions were more reticent to speak on these topics despite compensation. An overview of participants is provided in Table 1 (Ages suppressed to mitigate the risk of reidentification).

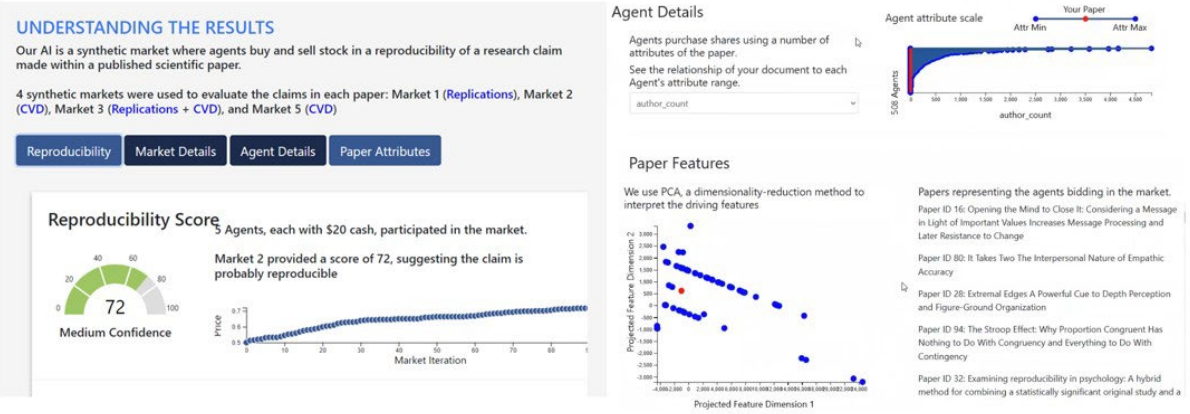

Figure 1. Screen capture from a prototype research assessment tool. Separate tabs allow the user to explore the AI and its functionality, including extracted features from the paper of interest and a subset of similar papers from the training dataset.

ID	Gender	University	Profession	Research field
P1	Female	Government	PhD	Social psychology
P2	Male	Government	Prof	Economics
P3	Male	Government	Assistant Prof	Mechanical engineering
P4	Male	Government	Assistant Prof	Electrical engineering
P5	Male	Government	Assistant Prof	Electrical engineering
P6	Male	Government	PhD	Machine learning
P7	Female	Government	PhD	Psychology
P8	Male	Private	Assistant Prof	Electrical engineering
P9	Female	Government	PhD	Clinical psychology
P10	Male	Government	PhD	Social psychology
P11	Male	Government	Associate Prof	Electrical engineering
P12	Male	Government	Associate Prof	Electrical engineering
P13	Male	Government	Assistant Prof	Electrical engineering
P14	Male	Private	Assistant Prof	Image processing
P15	Male	Government	Prof	Power engineering
P16	Male	Government	Prof	Labor economics
P17	Female	Government	PhD	Social psychology
P18	Male	Government	Assistant Prof	Electrical engineering
P19	Male	Government	Prof	Mechanical engineering

Table 1: Gender, institution, rank, and research area of interview participants.

Interview protocol

Semi-structured interviews (Sadek et al., 2023) were conducted virtually via Zoom video conferencing between January 2024 and April 2024. Interviews lasted between 40 to 55 minutes, depending on the length of participant responses to interview questions. All interviews were recorded and transcripts for analysis were generated through transcription capabilities native to Zoom.

We asked participants to share their thoughts on the state of reproducibility, replicability, and open science in India and within their research communities. During the interview, we demonstrated to participants a prototype AI tool to estimate the replicability of a given finding. They were asked to provide feedback on the AI tool in various aspects.

An exemplar AI for research assessment

To ground our discussions of AI in research assessment, we demonstrated a prototype replication-prediction tool to participants. The tool takes the form of an artificial prediction market, in which algorithmic agents (or '*bot traders*') buy and sell assets corresponding to the outcomes of a notional replication study of a given finding. The closing price of a '*will replicate*' asset serves as a proxy for the estimated probability that the finding is replicable. The inner workings of this prototype have been described in prior publications, and a demonstration video is publicly available. Here, we summarise its core components.

1) *Feature extraction.* The system extracts a comprehensive set of features from the paper and its metadata, including bibliometric, venue-related, author-related, statistical, and semantic information.
2) *Market model.* Algorithmic agents are represented as convex regions in feature space. Each agent is furnished with an initial cash allocation and trades assets ('*will replicate*' or '*will not replicate*') based on whether a finding falls within its region. Market dynamics are influenced by

hyperparameters such as liquidity constant, market duration, number of agents, and inter-arrival rates.

3) *Agent training.* At the conclusion of each market corresponding to a training data point, profitable agents are retained while unprofitable ones are removed. A subset of the most profitable agents undergo mutation and crossover via genetic algorithms to expand the agent pool.
4) *Explainability.* Results are displayed to users through an interface that visualises features considered by agents, assets purchased, and asset price movements over time. Figure 1 provides example screenshots of this interface.

Data analysis

Interview transcripts were studied using thematic analysis (Blandford et al., 2016). This type of qualitative data analysis involves in-depth reading of transcripts to identify patterns across the dataset and derive a set of themes related to the research questions. We used a collaborative and iterative coding process (Ding et al., 2022; Huang et al., 2020). The first author read the interview transcripts multiple times to get familiar with the data. Open coding was then performed to identify initial codes. These codes were then organised into themes pertinent to the primary research questions. Two authors met periodically to discuss the identified themes' meanings, similarities, and differences and their relevance to the research questions. Final decisions regarding the retention, removal, or reorganisation of these themes were made collectively by the author team.

Findings

State of reproducibility and replicability in India

All participants were familiar with the concepts of scientific reproducibility and replicability, but only a few (P7, P8, and P12) were aware of ongoing scientific dialogue around related challenges, i.e., the '*replication crisis*'. For example, P3 said: '*I have heard about reproducibility, but not the crisis part. The majority of the researchers in my lab care about reproducibility. There are some who do not, but that percentage is rare. -P3*'

During the discussion, *every* interview participant echoed the importance of scientific reproducibility and replication. They saw these as fundamental requirements for any scientific study and emphasised their role in upholding scientific integrity.

Researchers felt that these concerns were not adequately and openly discussed. P5 commented that this topic should receive more attention.

> I must congratulate you that it is a very wonderful topic that you are discussing. We don't discuss this topic much in the open.... We have forgotten the quality. So, just in the number of publications it is coming. It is not a healthy practice at all. -P5

Social, cultural, and institutional context around research integrity

Openness–as authors

During interviews, P5, P15, P16, and P18 mentioned without prompting that they try to adopt reproducible research practices in their own work and promote these practices in their lab. They try to make their code and data available to other researchers and for peer review processes.

> We do sometimes deposit the data set, even for very curated data that we create, even codes as well. In many cases, we deposit for public access, at least for journal access, so that the reviewers can look at it and make a decision on the basis of that. -P16

This, however, was not the case among all researchers. P13 gave one explanation for the hesitation to share datasets, namely the perception of an opportunity cost associated with the potential

utilisation of the dataset for additional experiments and future publications. '*The data that we have used for doing one work can be used for many other applications and maybe we can publish 3 more papers out of it.* -P13'

Openness–as readers

As consumers of others' research, participants reported challenges interpreting existing literature due to deficiencies in scholarly publishing practices, particularly insufficient information sharing. In computational research, this often manifests as a lack of code and data sharing, as P6 pointed out.

> There are some types of challenges that I face especially when the researchers have not given the code for their study. So, in some other cases, the challenges that we generally face is that the authors have used some kind of dataset, and maybe they have not mentioned how to get that dataset or to generate that dataset. -P6

Participants also reported insufficient detail around experiment design or methodology in publications. Researchers (P3, P4, P5, P6, P8, and P15) noted encountering significant challenges stemming from missing specifications where multiple essential parameters or experimental conditions were often omitted in publications. They also reported difficulty reproducing experiments with significant hardware requirements. P5 shared his experience with this challenge.

> So, a very common challenge is that in most of the papers, they don't describe the experimental setup very nicely. In many cases, people simply give up a schematic diagram. They don't even mention the model diagram and all other details. That becomes a real challenge for us. We are not able to clearly identify which equipment they have used, for which process. -P5

Pressures to publish

Our participants discussed the demand for novel findings as a major contributor to the reproducibility crisis (P9, P10, and P18). Because journals and conferences do not readily publish replication attempts, conducting replication studies may be perceived as less valuable or might even have adverse effects on their career.

> I feel that there is a kind of negative stigma associated with people who do replication studies. They are like, you cannot bring something new to the field of research. Therefore, even after hard work, no one values it. -P10

Participants reported pressure for high publication output. Adopting reproducible practices is a time-intensive endeavor, but incentives target publishing a greater number of articles rather than prioritising the quality of the work. As noted by P14, researchers with higher numbers of publications tend to reap rewards. '*In India, the scenario is that if someone has 20 papers and some academic contacts, they will get more preference. There is no measure of how to quantify their quality.* -P14'

Nevertheless, participants were optimistic that there is gradual but slow improvement. P16 acknowledged that more highly reputed journals have begun to adopt guidelines that require authors to share data and code.

Thoughts on AI-supported research assessment tools

During their interview, each participant observed the interviewer demonstrate a prototype AI replication prediction tool (Figure 1). The demonstration was used to seed broader conversation about participants' perceptions and suggestions for AI tools to assess published research findings.

Utility and design implications

All 19 participants expressed that they had not encountered a tool capable of predicting replicability and were excited to observe how it works. All participants expressed a desire to learn about the study and share their perspectives, recognising the significant relevance: '*I feel that the study is very exciting and challenging in its own way, and I personally feel that if this method is introduced, then it will be much easier for researchers to evaluate research papers. -P1*'

Participants perceived the tool to be useful. P11, an associate professor thinks this tool will be very helpful to the research community, noting it may take some time to adapt to the change. '*I feel maybe (something like this) will come soon. Maybe this will be almost a compulsory part of any research. I'm really feeling excited. Yes, Of course, it will be very helpful. -P11*'

Participants (P4, P13) proposed that AI for replication prediction could be particularly beneficial for researchers who lack the experience or resources to validate others' work, e.g., early career researchers (P4) and those lacking access to research facilities (P13). There are many universities in India where researchers don't have much funding to access expensive resources. Participants noted that such tools could expedite literature review (P5) and help to identify studies for further investigation (P6). One participant (P10) suggested that the tool's output could be integrated into the paper's metadata.

P11 and P12 suggested that tools for research assessment should be tailored to each domain. For experimental work, P12 suggested the software could generate two evaluations–one for the experimental design and another for analyses. This would help to differentiate between reproducibility or replicability of experiments vs. analyses. '*Maybe the analysis part is quite reproducible, but not the experimental part. So, if you can say we get different components. If you can add, I think that will be a good thing to do. -P12*'

Transparency, explainability and trust

Our interviews revealed a strong preference among participants (P2, P3, P6, P10, P12, P19) for a tool that adheres to open science and open-source best practices. Several participants highlighted the need for transparency with respect to the algorithm training process. P2 suggested tools for research assessment must make clear what features of the paper and its metadata are taken into consideration during evaluation.

Explainability was also a key point of discussion. P3 and P10 suggested that the tool should display not only a confidence score but also a brief explanation of the score in simple language. For example, P3 suggested the AI could provide information related to the availability of data and code for each paper. '*I think it is also necessary to point out the papers that are not reproducible, but it will be helpful if the tool can provide those papers having data missing. -P3*'

Trust was closely related to transparency and explainability of the scoring system. '*The explainability of any AI is required to build up trust. So, if you can make it more explainable, that means how this score is calculated. If you can explain that convincingly it will get more trust. -P12*'

While most researchers interviewed expressed great enthusiasm for the tool we shared and appreciated its utility, the majority struggled to place complete trust in any outcome generated by an AI model. With the exception of P2, P12, and P16, all participants expressed that they would have greater trust in systems facilitating human-AI collaboration (vs. fully autonomous systems) for research assessment. In particular, P5, P6, P11, and P18 stated that they trust AI in general but would prefer a hybrid human-AI system capable of integrating human feedback for this task. '*It will be very helpful if human intelligence and AI can work together, and if you can create such kind of interface. That will be really awesome. -P5*'

Lack of trust in fully AI systems stemmed from participants' awareness of their limitations and biases.

> So, the problem of all AI-based systems is that they have some inherent probability. That means we have to assume that they cannot be 100% accurate. And if the training data is not enough or the training is not accurate. They can provide drastically wrong results. That is the problem with these AI-based systems. But on average their performance is good. So, therefore, if I try this kind of system for so many databases, I believe that it will give you very good results. -P12

Similarly, several participants conveyed that they would not unconditionally trust an AI's output. They would assess reproducibility and replicability with their own expertise but that does not mean that they would not also use available tools. P15, for instance, stating that he would personally verify the results but would use the AI as an additional input.

> I wouldn't rely on software totally. Rather, I would judge my intuition or use my intuition to check the veracity of the results. I mean compound to what I just said is that then does it mean that I will not be using it? No, I will use it to see the output. -P15

Participants also expressed that trust would need to be earned. If AI tools for research assessment work well for researchers, trust will gradually build. In addition, P5 suggested that trust in any tool is inherently intertwined with the research team responsible for its development.

Features to be extracted from papers and metadata

Experience was mentioned by many researchers as important for estimating the quality of research, but it comes with time and practice. We asked participants to share some of the tangible signals they look for within a paper or its metadata to estimate the reproducibility, replicability, and general credibility of work in their area. Ultimately, these signals may serve as input features to future algorithmic systems trained to estimate confidence in published findings.

Journal and author reputation

Journal impact factor, author, and paper citation counts, play an outsized role in perceptions of credibility, but our participants also noted their flaws (P11, P4). '*I think the journal impact factor plays an important role in judging the credibility of those works but I don't look for citation count and I think it can mislead the researchers. -P11*'

According to P2, researchers value journals that have high impact factors but that the calculation of impact factors depends on many things and varies across domains. P2 also noticed that people tend to cite papers that are (already) well-cited without much justification or knowledge about them.

On the other hand, P1 shared that journal impact factor and authors' reputations are both very important to her own paper evaluations. She believes that highly renowned authors are more trustworthy.

> I'll see the journals of high impact factors; when you go through those journals you will get some good ideas about how the paper is presented. You know impact factor is the one by which people get trust. Obviously, you know some authors over there are highly rated because they have some seriously good insights that will help upcoming researchers like us. -P1

P15 mentioned he believes that reputable journals have a high standard of peer review, therefore people tend to believe those results are reliable.

Methodology and study design

Methodology and study design are important considerations when estimating the credibility of reported findings. P6, a graduate student in machine learning, commented he always reads through the methodology carefully.

> One important criterion would definitely be going through the methodology section of any paper during evaluation. That would be one thing if I am clear about how they have collected the data and made the design procedure, I will consider it. -P6

P4 also mentioned the importance of methodology. P7, P10, and P14 mentioned the importance of sample size and validation of methods and reported outcomes, particularly when the methodology involves many hyper-parameters and other environmental details that should be clearly described. P11, whose research involves hardware testing, mentioned other factors like related work and experimental setup as important indicators of credibility in his domain. *I search papers based on my experience and see how logical is the overall paper in terms of the literature, survey, methodology, materials, experimental setup data, etc.* -P11

Open data and code

Open sharing of code and data plays a foundational role in practical reproduction and replication of study findings. Given the nature of their research, participants from engineering fields emphasised the importance of open code. For instance, P14, a researcher in the area of image processing, routinely seeks the source code when evaluating any paper and attempts to execute it for validation purposes.

> While doing the paper evaluation, I search for the source code. If the source code has already been given by the authors, then I try to run those and evaluate the results. I also use the code with our new data and see how that performs. -P14

P18 discussed the importance of data sharing.

> Say, I'm doing primary survey-based research. Okay, then, I've got the data. I have to share the data with others in my publication and reported results. Then only others can reproduce my results, of course; otherwise, without the data, how can they reproduce my results? -P18

Discussion

We explore opportunities and challenges for AI in the space of research assessment, particularly design opportunities for replication prediction. Our study centers on the perspectives of researchers in India, broadening discourse as a unique contribution to research integrity in the Global South. The rapid proliferation of research has long created noise for researchers, but it is brought into stark relief in the era of generative AI. In this context, technologies to support automated evaluation of published work along axes of reproducibility, replicability, and robustness are critically important. These technologies will presumably be driven by advances in NLP, ML, information retrieval, computer vision, and similar.

Our work highlights the fact that technologies will be situated amongst complex social, cultural, and institutional processes of scientific practice. AI tools for research assessment will process research outputs that are available. This can help junior research scholar with their literature review process. They will require better documentation, preservation, and open-sharing practices. Open science has gained considerable attention in recent years, with India actively participating in global initiatives (Fernandez, 2006; Scaria & Ray, 2018) and taking some initial steps to promote open science practices (Hanief Bhat, 2014; Nazim et al., 2023). The Indian government and various institutions have begun supporting and promoting open access (OA) publication in the country. One notable example is the Shodhganga repository, which is a reservoir of Indian theses and

dissertations. Managed by the Information and Library Network Centre (INFLIBNET) under the University Grants Commission (UGC), Shodhganga promotes open access to Indian scholarly works, including Ph.D. theses submitted to universities in India (Sivakumaren, 2015). But the private universities are still far away from open science and open culture.

In this study, we observed that researchers have ethical concerns about using AI tools for replication prediction and need more explainability. From the discussion, it came out which features are important for these kinds of systems and how they can integrate these systems into their research. Despite this, our findings suggest that the successful implementation of open science practices remains a challenge in India (Nazim et al., 2023). One major issue is the challenge Indian researchers may face in publishing if open access fees are high, since many institutions lack the funding to cover these costs. Additionally, the very competitive nature of funding in India and the limited availability of grants incentivise researchers to maximise the use of their data for multiple publications rather than sharing it openly after a single publication because of the publish and perish culture (Bello et al., 2023; Van Dalen & Henkens, 2012). This practice is driven by a broader academic culture that tends to value the quantity of publications over quality. Our participants expressed enthusiasm that this study would serve as a voice for their challenges and act as a catalyst to improve scientific research in India.

Design principles for scholarly tools

Well-designed AI tools with explainability

Our findings underscore the potential value of AI research assessment tools that can offer context to researchers about published findings, and aspirationally, estimate research reproducibility, replicability, generalisability, and beyond. Our study finds that these kinds of tools would assist researchers, especially early career researchers, during literature review when they evaluate other papers, specifically for junior scholars who have very little experience. Previous work suggests that there are tools for literature review, but not with advanced features like this (Boeker et al., 2013). Our participants also expressed that such tools would help them think about reproducibility and replicability from the very beginning of each study they design.

Features from papers and metadata

Our participants enumerated several specifications for the design and development of such technologies. They suggested that features to be considered when evaluating a published finding should include domain-specific knowledge. Different types of research domains evaluate papers based on different criteria. Domain specific features can be used for better results and query suggestions to users (White et al., 2009). These tools should have a feature to filter out the results based on their research domains. However, all the researchers mentioned theoretical basis, sample size, and statistical power as important features to be integrated. We note that social scientists and engineers in our study differed somewhat in their perspectives, e.g., on the importance of journal reputation and data sharing. These distinctions highlight the need for future AI-driven technologies in this space to support nuance across disciplines. Most of the researchers don't look for preregistration when they evaluate the credibility of any research article. Therefore, these tools should represent all the features clearly to the users based on what it got trained which can help the users' final decision.

Transparency, bias, and fairness

Our participants widely agreed that they would be more inclined to trust AI tools for research assessment if they made transparent their training data, features, and algorithms (de Brito Duarte et al., 2023). They stressed the importance of model explainability for building trust and fostering adoption. The current tool does not provide detailed explanations for the decision-making process and the training process. AI technologies for research assessment should be designed and deployed in a manner that promotes fairness and minimises bias (Chen et al., 2023). This involves careful consideration of training data, algorithmic design, and validation methods to ensure

equitable outcomes across diverse populations. For example, should assessment algorithms consider authors' institutional affiliations, rank, or funding? These are questions that warrant deep consideration by the research community and stakeholders.

Human-AI technologies

As research assessment is a complicated and sensitive task, our study suggests that blind trust in AI is neither realistic nor prudent. The majority of researchers we interviewed expressed a preference for hybrid systems that could integrate inputs from both human experts and AI, which was eventually observed by many other researchers (Jiang et al., 2023; Memmert & Bittner, 2022). Human experts bring domain knowledge, intuition, and contextual understanding to the decision-making process, while AI algorithms excel at processing large volumes of data, identifying patterns, and making predictions. By combining the strengths of both, the hope is that hybrid systems can generate more accurate and reliable insights than either humans or AI alone. Yet, how these technologies might be designed and deployed in practice is an open question.

Policy directions for institutional support

Repositories and digital libraries

Platforms like Shodhganga, OSF, and Zenodo can embed replication-prediction scores into publication metadata, allowing researchers to access credibility cues during literature search. Policy frameworks should encourage standardised documentation of openness, reproducibility, and replicability.

Universities and libraries

Institutions should move beyond workshops and integrate mandatory graduate level coursework in data curation, open methods, and responsible AI into their curriculum. Such formalised training ensures that transparency and reproducibility are embedded early in scholarly practice rather than treated as optional. Alongside education, universities should reform reward structures by explicitly recognising open data and code contributions, as well as community resources such as datasets and tutorials, in tenure and promotion guidelines, placing them on par with traditional research outputs. Finally, libraries, specifically in India, host shared infrastructure such as institutional GitLab instances and mediate between researchers and repositories to provide sustainable, reliable support for open science.

Funding agencies

Granting bodies should offset open access fees, which remain prohibitive for many researchers in India and the Global South. Funding mandates for data and code sharing should be paired with infrastructure and training support. Investment in culturally sensitive AI research assessment tools can further advance equity.

Journals and conferences

Editorial policies can support replication studies by providing dedicated publication tracks and implementing reproducibility checklists or transparency badges. Conferences can serve as testbeds for the responsible integration of AI-supported tools into peer review and scholarly communication.

Limitations

This study offers valuable insights into researchers' perspectives on AI-assisted replication prediction, particularly within the Indian academic context. The findings are based on a qualitative sample from select disciplines and institutions, which may not fully represent the diversity of research practices across globally. While the design recommendations are grounded in participant input, they remain exploratory and have not yet been tested in applied settings. Using the tool in actual research settings would provide a more comprehensive understanding of its impact on the

research life cycle. We hope that this work will catalyse similar studies focused on other countries and broad recognition of the importance of cultural context in the discussion of these issues.

Conclusion

We study the opportunities and challenges for AI-supported research assessment, with specific focus on India. We acknowledge that the sheer volume of scientific output makes it impossible to reproduce and replicate every finding in the literature. Accordingly, we explore researchers' openness to AI-driven technologies to support the estimation of reproducibility and replicability. We outline design requirements for these systems, in particular, finding that hybrid human-AI technologies are viewed as most useful and trustworthy for now. While there is consensus regarding the utility of AI-supported tools, we find differences in opinion regarding the training, design, and envisioned implementation of these technologies. Moving forward, addressing concerns related to transparency, explainability, and human-AI collaboration will be foundational for trust and adoption within the research community. Our findings highlight the critical importance of context for harnessing the potential of AI in the space of research assessment. While there is general recognition of the value of openness and reproducibility, researchers lack resources and incentives to meticulously document research processes and share them openly. Some researchers choose to engage in these practices driven by their personal ethics. Ultimately, resources, incentives, and ethics are deeply social and cultural, scaffolded by institutional structures. Our work suggests that, as such, it is critical that the research community engage with a full and representative set of research stakeholders throughout the globe. This is an opportune moment to do so, given the nascent use of AI technologies in this space.

Acknowledgements

We acknowledge support by DARPA W911NF-19-2- 0272. This work does not necessarily reflect the position or policy of DARPA, and no official endorsement should be inferred.

About the authors

Tatiana Chakravorti is a PhD candidate in Informatics at Pennsylvania State University College of Information Sciences and Technology. Her research interests centre on hybrid human-AI systems, information integrity, and machine learning applications for smart technologies. She can be contacted at tfc5416@psu.edu.

Chuhao Wu is a Strategic Assessment Librarian at Clemson University Libraries. He completed his PhD in Informatics at Pennsylvania State University. His research focuses on human-computer interaction, educational technologies, and socio-technical aspects of AI. He can be contacted at chuhaow@clemson.edu.

Sai Koneru is a PhD candidate in the College of Information Sciences and Technology at Pennsylvania State University. His research interests include scientific text processing, information integrity, and meta-science. He can be contacted at sdk96@psu.edu.

Sarah Rajtmajer is an Associate Professor in the College of Information Sciences and Technology at Pennsylvania State University. Dr. Rajtmajer holds a PhD in Mathematics from the University of Zagreb. Her research integrates machine learning, AI, and hybrid human-AI systems to study how information encodes values such as accuracy, objectivity, and privacy. Her work spans computational social science, ethical AI, and online socio-technical systems. She can be contacted smr48@psu.edu.

References

Ahmad, M. A., Teredesai, A., & Eckert, C. (2020). Fairness, accountability, transparency in ai at scale: Lessons from national programs. Proceedings of the 2020 conference on fairness, accountability, and transparency, 690–690.

Altmejd, A., Dreber, A., Forsell, E., Huber, J., Imai, T., Johannesson, M., Kirchler, M., Nave, G., & Camerer, C. (2019). Predicting the replicability of social science lab experiments. PloS one, 14(12), e0225826.

Anderson, M. S., Ronning, E. A., Vries, R. D., & Martinson, B. C. (2010). Extending the mertonian norms: Scientists' subscription to norms of research. The Journal of higher education, 81(3), 366–393.

Bajpai, V., Kühlewind, M., Ott, J., Schönwälder, J., Sperotto, A., & Trammell, B. (2017). Challenges with reproducibility. Proceedings of the Reproducibility Workshop, 1–4.

Baker, M. (2016a). 1,500 scientists lift the lid on reproducibility. Nature, 533(7604). Baker, M. (2016b). Reproducibility crisis. Nature, 533(26), 353–66.

Bánáti, A., Kacsuk, P., & Kozlovszky, M. (2015). Four level provenance support to achieve portable reproducibility of scientific workflows. 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 241–244.

Bello, S. A., Azubuike, F. C., & Akande, O. A. (2023). Reputation disparity in teaching and research productivity and rewards in the context of consequences of institutionalIsation of publish or perish culture in academia. Higher Education Quarterly, 77(3), 574–584.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., et al. (2018). Redefine statistical significance. Nature human behaviour, 2(1), 6–10.

Blandford, A., Furniss, D., & Makri, S. (2016). Qualitative hci research: Going behind the scenes. Morgan & Claypool Publishers.

Boeker, M., Vach, W., & Motschall, E. (2013). Google scholar as replacement for systematic literature searches: Good relative recall and precision are not enough. BMC medical research methodology, 13, 1–12.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., et al. (2016). Evaluating replicability of laboratory experiments in economics. Science, 351(6280), 1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. Nature Human Behaviour, 2(9), 637–644.

Chakravorti, T., Dileep Koneru, S., & Rajtmajer, S. (2024). Reproducibility, replicability, and transparency in research: What 430 professors think in universities across the usa and india. arXiv e-prints, arXiv–2402.

Chard, K., Pruyne, J., Blaiszik, B., Ananthakrishnan, R., Tuecke, S., & Foster, I. (2015). Globus data publication as a service: Lowering barriers to reproducible science. 2015 IEEE 11th International Conference on e-Science, 401–410.

Chen, P., Wu, L., & Wang, L. (2023). Ai fairness in data management and analytics: A review on challenges, methodologies and applications. Applied Sciences, 13(18), 10258.

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Collberg, C., & Proebsting, T. A. (2016). Repeatability in computer systems research. Commun. ACM, 59(3), 62–69. https://doi.org/10.1145/2812803

Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., Beebe, J., Berniūnas, R., Boudesseul, J., Colombo, M., et al. (2021). Estimating the reproducibility of experimental philosophy. Review of Philosophy and Psychology, 12(1), 9– 44.

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., Orben, A., Parsons, S., & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. Zeitschrift für Psychologie.

Dacrema, M. F., Boglio, S., Cremonesi, P., & Jannach, D. (2021). A troubling analysis of reproducibility and progress in recommender systems research. ACM Trans. Inf. Syst., 39(2), 20:1–20:49. https://doi.org/10.1145/3434185

de Brito Duarte, R., Correia, F., Arriaga, P., & Paiva, A. (2023). Ai trust: Can explainable ai enhance warranted trust? Human Behavior and Emerging Technologies, 2023(1), 4637678.

Ding, X., Kou, Y., Xu, Y., & Zhang, P. (2022). 'as uploaders, we have the responsibility': Individualised professionalIsation of bilibili uploaders. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–14.

Dutta, M., Ramasubramanian, S., Barrett, M., Elers, C., Sarwatay, D., Raghunath, P., Kaur, S., Dutta, D., Jayan, P., Rahman, M., et al. (2021). Decolonising open science: Southern interventions. Journal of communication, 71(5), 803–826.

Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. Elife, 3, e04333.

Eyert, F., & Lopez, P. (2023). Rethinking transparency as a communicative constellation. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 444–454.

Feger, S. S., Dallmeier-Tiessen, S., Schmidt, A., & Woźniak, P. W. (2019). Designing for reproducibility: A qualitative study of challenges and opportunities in high energy physics. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–14.

Fernandez, L. (2006). Open access initiatives in india-an evaluation. Partnership: The Canadian Journal of Library and Information Practice and Research, 1(1). Fidler, F., & Wilcox, J. (2018). Reproducibility of scientific results.

Freese, J., Rauf, T., & Voelkel, J. G. (2022). Advances in transparency and reproducibility in the social sciences. Social Science Research, 107, 102770.

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. Annals of the New York Academy of Sciences, 1396(1), 5–18.

Gohel, P., Singh, P., & Mohanty, M. (2021). Explainable ai: Current status and future directions. arXiv preprint arXiv:2107.07045.

Gunzer, F., Jantscher, M., Hassler, E. M., Kau, T., & Reishofer, G. (2022). Reproducibility of artificial intelligence models in computed tomography of the head: A quantitative analysis. Insights into Imaging, 13(1), 1–8.

Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Shraddha, T., Kusko, R., Sansone, S.-A., Tong, W., Wolfinger, R. D., Mason, C. E., Jones, W., Dopazo, J.,

Haim, A., Baxter, C., Gyurcsan, R., Shaw, S. T., & Heffernan, N. T. (2023). How to open science: Analysing the open science statement compliance of the learning@ scale conference. Proceedings of the Tenth ACM Conference on Learning@ Scale, 174–182.

Hanief Bhat, M. (2014). Exploring research data in indian institutional repositories. Program, 48(2), 206–216.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. PLoS biology, 13(3), e1002106.

Huang, X., Vitak, J., & Tausczik, Y. (2020). 'You don't have to know my past': How wechat moments users manage their evolving self-presentation. Proceedings of the 2020 CHI Conference on Human Factors in Computing systems, 1–13.

Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. Science, 359(6377), 725–726. https://doi.org/10.1126/science.359.6377.725

Iyer, R., Li, Y., Li, H., Lewis, M., Sundar, R., & Sycara, K. (2018). Transparency and explanation in deep reinforcement learning neural networks. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 144–150.

Jiang, N., Liu, X., Liu, H., Lim, E. T. K., Tan, C.-W., & Gu, J. (2023). Beyond ai-powered context-aware services: The role of human–ai collaboration. Industrial Management & Data Systems, 123(11), 2771–2802.

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machinelearning-based science. Patterns, 4(9).

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. Science, 355(6325), 584–585.

López-Nicolás, R., López-López, J. A., Rubio-Aparicio, M., & Sánchez-Meca, J. (2022).

A meta-review of transparency and reproducibility-related reporting prac-

tices in published meta-analyses on clinical psychological interventions (2000–2020). Behavior research methods, 54(1), 334–349.

Lu, J., Lee, D., Kim, T. W., & Danks, D. (2019). Good explanation for algorithmic transparency. Available at SSRN 3503603.

Marshall, I. J., & Wallace, B. C. (2019). Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. Systematic reviews, 8, 1–10.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? what does 'failure to replicate' really mean? American Psychologist, 70(6), 487.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., et al. (2016). How open science helps researchers succeed. elife, 5, e16800.

Mede, N. G., Schäfer, M. S., Ziegler, R., & Weißkopf, M. (2021). The 'replication crisis' in the public eye: Germans' awareness and perceptions of the (ir) reproducibility of scientific research. Public Understanding of Science, 30(1), 91– 102.

Memmert, L., & Bittner, E. (2022). Complex problem solving through human-ai collaboration: Literature review on research contexts.

Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis.

National Academies of Sciences, Engineering, and Medicine et al. (2019). Reproducibility and replicability in science. National Academies Press.

Nazim, M., Bhardwaj, R. K., Agrawal, A., & Bano, A. (2023). Open access publishing in india: Trends and policy perspectives. Global Knowledge, Memory, and Communication, 72(4/5), 437–451.

Niksirat, K. S., Goswami, L., Tyler, J., Silacci, A., Aliyu, S., Aebli, A., Wacharamanotham, C., Cherubini, M., et al. (2023). Changes in research ethics, openness, and transparency in empirical studies between chi 2017 and chi 2022.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., et al. (2015). Promoting an open research culture. Science, 348(6242), 1422–1425.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Breckler, S., Buck, S., Chambers, C., Chin, G., Christensen, G., et al. (2016). Transparency and openness promotion (top) guidelines.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences, 115(11), 2600–2606.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Almenberg, A. D., Fidler, F., Hilgard, J., Kline, M., Nuijten, M. B., et al. (2021). Replicability, robustness, and reproducibility in psychological science.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: Ii. restructuring incentives and practices to promote truth over publishability. Perspectives on Psychological Science, 7(6), 615–631.

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. Advances in Methods and Practices in Psychological Science, 3(2), 229–237.

Patil, P., Peng, R. D., & Leek, J. T. (2016). A statistical definition for reproducibility and replicability. BioRxiv, 066803.

Pawel, S., & Held, L. (2020). Probabilistic forecasting of replication studies. PloS one, 15(4), e0231416.

Pimentel, J. F., Murta, L., Braganholo, V., & Freire, J. (2019). A large-scale study about quality and reproducibility of jupyter notebooks. 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR), 507–517. https: //doi.org/10.1109/MSR.2019.00077

Polanin, J. R., Hennessy, E. A., & Tsuji, S. (2020). Transparency and reproducibility of meta-analyses in psychology: A meta-review. Perspectives on Psychological Science, 15(4), 1026–1041.

Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data cards: Purposeful and transparent dataset documentation for responsible ai. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 1776– 1826.

Raff, E. (2019). A step toward quantifying independently reproducible machine learning research. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, & R. Garnett (Eds.), Advances in neural information processing systems 32: Annual conference on neural information processing systems 2019, neurips 2019, december 8-14, 2019, vancouver, bc, canada (pp. 5486–5496).

Rajtmajer, S., Griffin, C., Wu, J., Fraleigh, R., Balaji, L., Squicciarini, A., Kwasnica, A., Pennock, D., McLaughlin, M., Fritton, T., et al. (2022). A synthetic prediction market for estimating confidence in published work. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 13218–13220.

Rowe, A. (2023). Recommendations to improve use and reporting of statistics in animal experiments. Laboratory Animals, 57(3), 224–235.

Sadek, M., Calvo, R. A., & Mougenot, C. (2023). Trends, challenges, and processes in conversational agent design: Exploring practitioners' views through semistructured interviews. Proceedings of the 5th International Conference on Conversational User Interfaces, 1–10.

Samuel, S., & König-Ries, B. (2022). A collaborative semantic-based provenance management platform for reproducibility. PeerJ Computer Science, 8, e921.

Santana-Perez, I., Pérez-Hernández, M. S., et al. (2015). Towards reproducibility in scientific workflows: An infrastructure-based approach. Scientific Programming, 2015.

Scaria, A. G., & Ray, S. (2018). Open science india report.

Schönbrodt, F., & Mellor, D. (2018). Academic job offers that mentioned open science. open science framework.

Schooler, J. W. (2014). Metascience could rescue the 'replication crisis'. Nature, 515(7525), 9–9.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. Annual review of psychology, 69, 487–510.

Sivakumaren, K. (2015). Electronic thesis and dissertations (etds) by indian universities in shodhganga project: A study. Journal of advances in library and information science, 4(1), 62–66.

Sokol, K. (2019). Fairness, accountability, and transparency in artificial intelligence: A case study of logical predictive models. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 541–542.

Stodden, V., Leisch, F., & Peng, R. D. (2014). Implementing reproducible research. CRC Press.

Stodden, V., & Miguez, S. (2013). Best practices for computational science: Software infrastructure and environments for reproducible and extensible research. Available at SSRN 2322276.

Thakkar, D., Ismail, A., Kumar, P., Hanna, A., Sambasivan, N., & Kumar, N. (2022). When is machine learning data good?: Valuing in public health datafication. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–16.

Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., McCarthy, R. J., Riegelman, A., Silberzahn, R., & Nosek, B. A. (2019). Scientific utopia iii: Crowdsourcing science. Perspectives on Psychological Science, 14(5), 711–733.

Van Dalen, H. P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. Journal of the American Society for Information Science and Technology, 63(7), 1282–1293.

Varanasi, R. A., Siddarth, D., Seshadri, V., Bali, K., & Vashistha, A. (2022). Feeling proud, feeling embarrassed: Experiences of low-income women with crowd work. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–18.

Vilhuber, L. (2020). Reproducibility and replicability in economics. Harvard Data Science Review, 2(4), 1–39.

White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterising the influence of domain expertise on web search behavior. Proceedings of the second ACM international conference on web search and data mining, 132–141.

Willis, C., & Stodden, V. (2020). Trust but verify: How to leverage policies, workflows, and infrastructure to ensure computational reproducibility in publication.

Yang, Y., Youyou, W., & Uzzi, B. (2020). Estimating the deep replicability of scientific findings using human and artificial intelligence. Proceedings of the National Academy of Sciences, 117(20), 10762–10768.