Trustworthiness of and Trust in Research Infrastructures: How to Foster Quality & Accountability in Data Driven Research

Contributors:
Andreas Rauber, Technische Universität Wien & SBA-Research
Satoshi Oyama, Nagoya City University
Kathrin Grosse, IBM Research Zurich
Balázs E. Pataki, HUN-REN SZTAKI
András Micsik, HUN-REN SZTAKI
László Kovács, HUN-REN SZTAKI
Katharina Flicker, Technische Universität Wien & SBA-Research

Abstract

With the increasing complexity of research processes and the growing importance of various kinds of research infrastructure services in supporting research, the possibility of full end-to-end control and verification is decreasing. This raises challenges in terms of trustworthiness and, ultimately, trust in the data, tools, services and results obtained. The purpose of this document is to analyze the challenges emerging, and to identify gaps requiring more in-depth studies to ensure that researchers can perform high-quality research. As traditional indicators of trustworthiness are losing their indicative value, novel mechanisms will need to be identified and their documentation supported by the respective infrastructures. This is essential to ensure that the investments flowing into the design, creation and operation of global research infrastructures are sustainable, allowing researchers to have confidence in the results they produce, and society to trust scientific insights.

Introduction

Research challenges and data driven research across all disciplines are increasingly interdisciplinary and affect actual research practices as they too become more complex. First, research data, code, machinery collected, or built in one setting may be used in various research contexts. Such research practices might lead to new requirements such as expertise from different disciplines due to the lack of required knowledge. In other words, research is becoming more interdisciplinary.

Second, the number of institutions and individuals participating globally in research is growing dramatically, making it harder to know which ones are deserving of a good reputation in terms of high-quality and trustworthiness, all while the value of traditional indicators (such as relying on personal acquaintances, on the reputation of laboratories or institutions, on the use of a few prominent, well-known software frameworks or libraries, or on peer-reviews) of trust and trustworthiness are eroding.

Third, the speed of scientific knowledge production is increasing (e.g. pre-prints), leading to less time for quality checks, the maturation of knowledge, or to evaluate the entire research process, all while putting researchers under pressure to use the most recent innovations (the most recent LLM/AI, the most recent data set released), or to provide high frequency updates to code, fixing bugs and adding functionality, presumably improving performance in specific settings. Consequently, the risk of 'tuning' results or exaggerating the relevance of findings increases [2], [6].

Fourth, contexts in which data, code and entire processing pipelines are being (re-)used vary as data might be collected for one setting but are used in another, code repurposed for other tasks, with code itself becoming increasingly distributed and dynamic due to the large number of external dependencies. In short, many steps of such research processes are out of a researchers' control meaning that it is not possible to run and check everything in detail anymore because of, e.g. a lack of skills, time, money and necessary information.

Against this background, Research Infrastructures (RIs) providing access to the various elements of these increasingly complex research processes are gaining importance in terms of, for example, providing data, code, workflow engines, compute resources, results/research outputs, or result discussions (e.g. open reviews). The increasing importance of RIs lead to two developments: Significant resources are invested in the building of new RIs on the institutional (e.g. DataRepositories integrated with Jupyter Hubs), discipline-specific (e.g. ERICs), or national level (e.g. NFDI), or even in larger federations (e.g. EOSC, GAIA-X). Similar initiatives are also being developed in Japan, including institutional-level deployments of research data platforms such as NII's GakuNin RDM, discipline-specific repositories like the National Bioscience Database Center (NBDC) for life sciences, and national-level coordination through the Open Science Framework led by MEXT and JST. While Japan currently lacks large-scale federated structures comparable to EOSC, there is increasing emphasis on interoperability and metadata standardization in alignment with international practices.

Second, this comes with plenty of challenges: as there is no full control and no possibility for complete end-to-end checking, researchers basically need to trust in the components, data, and compute environments to be (re-)used without being able to verify them in full depth. Last, but not least, the concept of research infrastructure and the services provided by them is evolving, ranging from storage- compute- and networking infrastructure, via data and code provisioning to AI systems and services on the technical level, combined with authentication information on natural and legal persons, ethical, legal and financial frameworks and associated services. All of these have an impact on the research performed and its perception by researchers and society.

This in return leads to different types of questions coming up with respect to trust, trustworthiness and accountability:

- How can researchers decide whether (or to what extent) they can trust what they are working with in terms of data, services, tools and RIs? Why trust a specific component and not another? Why are researchers reusing this or that while ignoring other elements, or rather what makes data, services, tools and RIs trustworthy?
- Hence, for RI design: which information, characteristics, aspects, do RIs need to collect, expose and consider in their design and operation, so that they act as trustworthy sources and basis for producing high-quality research?
- What is trust supposed to be in the context of scientific knowledge production (that is rational and evidence based) as it cannot be expected to be universal and "blind"? To what extent can researchers trust their own research outputs? To what extent can and should researchers be held accountable for their results, or rather how can researchers accept accountability for their results?

To address these challenges we already see many attempts to develop new indicators of trustworthiness such as social-media inspired likes, download/fork counts in code repositories, reproducibility badges, automatic provenance documentation, persistent identifiers for humans (ORCID) and digital artifacts, rankings of institutions, open review processes, community review / agreement processes, etc. All these initiatives, however, are ad-hoc, and they were not designed to be trustworthy for researchers specifically. Hence, they are neither coordinated, nor supported by foundational analysis of needs and impact. To the best of our knowledge, there is no work on to which degree they can actually serve and be accepted by researchers as suitable indicators for trustworthiness. Many of them do not even provide any safeguards against "optimization", i.e. manipulation.

We therefore need a structured approach to derive aspects influencing perceptions of trustworthiness as well as indicators facilitating trust in such aspects. Such an approach is needed to understand how the perception in terms of strength, importance and robustness of these indicators differ across cultures, disciplines and seniority levels of researchers in order to provide guidance for all the RIs being developed: which data, metadata, provenance documentation, certification and assertions on data, code, institutions, humans etc. need to be captured and exhibited to ensure researchers find the structures and information required by them to make the personal decision whether they are able to trust in the building blocks of their own research as well as in the outcomes and insights they produce and consequently are able and willing to accept accountability.

In this context it is important to note that the quality of any artifact and hence its usefulness in a specific research setting is not an absolute characteristic that can be determined and ranked upfront: quality is frequently defined as "fitness for purpose" [10], hence the same piece of code or data, the same level of knowledge of a human expert may be perfect for one setting whereas it might be useless for another.

Against this background, this paper explores the concepts of "trust" as well as "trustworthiness" (as they are a subject in philosophical debate [12]) and moves on to discuss the different flavours and roles of RIs with a focus on the technical and human aspects of RIs, while also covering organisational, legal and financial aspects in section 2.

Section 3 provides early-stage results based on jointly organized workshops of the European Research Council for Informatics and Mathematics (ERCIM) and Japan Science and

Technology Agency (JST) as well as a survey, an interview series and informal discussions that were enabled by ASEA-UNINET, EOSC Focus and the EOSC Support Office Austria Working Group Researcher Engagement in Austria to help shape the discussion.

Based on that, we not only summarize the results, but also identify the work needed to come up with the insights needed to provide such guidance to RI developers and users and draw recommendations for the building of infrastructures from that in section 4.

2. Concepts

2.1 Trust & Trustworthiness

The concepts of trust and trustworthiness have long been subject to philosophical debate, often exploring various dimensions such as the nature of their concepts, their epistemologies, their values, or the kind of attitudes that trust is [12]. In the context of scientific knowledge production, we are, however, looking for types of warranted trust and trustworthiness because scientific knowledge production should not be “blind” but evidence-based and rational instead. Hereafter, the discussion focuses on when trust might be warranted and concludes with defining both trust and trustworthiness in a way that is applicable in the context of research.

Trust is warranted when it is “justified”, “well-grounded” and “plausible”. It is plausible when the trustor has reason (that is, evidence) to believe that the trustee is indeed trustworthy. It is well-grounded, when the trustee actually is trustworthy (which makes contemplating the nature of trustworthiness key if one strives to understand trust), while its justification may depend on either the epistemology of trust, or its value. The epistemology of trust matters when trustworthiness cannot be taken for granted. It deals with questions about how to trust well, or when to trust, while considering the value of trust means to question the point of trusting, or rather why trustors should trust at all. The value of trust may be intrinsic, but can be beneficial to society in general as it makes cooperation possible and facilitates relationships of all sorts [12].

Even though there is no agreement on whether trust can be rational indeed, it should be noted that philosophers who do think so, do not necessarily see eye to eye to the extent to which trust can be rational. One point of view, however, suggests that trust can indeed be rational and evidence-based. Such an internalist perspective states that reasons grounded in evidence are needed for the trustor to actually trust, so that both reason and evidence may internally justify trust [12].

The concept of trustworthiness is hard to grasp since it is being discussed quite controversially with a lot of different lines of argumentation to look at. They usually are associated with interpersonal trust and emphasize e.g. the relationship between trustor and trustee, motivation, goodwill, or virtue. In other words, the nature of trustworthiness is not quite clear. In addition, there is little agreement on its definition [12]. On top of that we are looking for a simple definition that allows us to think of “things” such as data, or infrastructures, rather than people, as trustworthy.

A practical approach to this may be based on three aspects [11]: First, trustworthiness is a trait. Second, there is a difference between general and specific trustworthiness. Third,

trustworthiness means to meet reasonable and appropriate expectations. The latter leads to questions of what makes them both. The second allows us to think of trustworthiness in a specific context - that is scientific knowledge production - rather than generalize it. The first one permits us to think of trustworthiness as a trait, an attribute, a characteristic of someone or something. Against this background, we can start looking for traits of data, infrastructures, etc. that are indicative of trust in them.

2.2 Research Infrastructures (RIs)

As [13] point out, defining infrastructures is a difficult matter. A simplified - and thus flawed - definition may state that infrastructures support and enable activities that we are really doing, while being built on prior work in terms of e.g. buildings, or standards. Hence, it refers to systems, technologies, organisations and built artifacts that do not need to be reconsidered because they already exist. In addition, infrastructures tend to be (made) invisible, although they are not a neutral background enabling different sets of activities. Rather, they hold values, or permit specific relations, while blocking others.

In Science and Technology Studies (STS) infrastructures are therefore conceptualized as a bundle of many heterogeneous things (for example, standards, technological objects, administrative procedures) that involve technology as well as organisational work. In other words, infrastructure is being treated relationally [13] .

Exactly because infrastructure is a highly relational concept, [14] defines infrastructure as having specific properties:

- Infrastructure is **embedded**, meaning that it is sunk into and inside of other structures.
- Infrastructure is **transparent** meaning that it does not need to be reinvented for each and every single task. However, it supports all tasks performed invisibly.
- **Reach & Scope:** Infrastructure is not limited to a single event or one-site practice.
- The use of an infrastructure is **learned as part of a membership** or a community of practice. It is taken for granted by members, while outsiders or strangers perceive infrastructure as something to learn about.
- Infrastructure shapes and is shaped with the **conventions of practice** by a community of practice.
- **Standards are embodied** in that infrastructures are plugging into other infrastructures and tools in a standardized mode.
- Infrastructure is **built on an installed base** inheriting both strengths and limitations.
- Infrastructure **becomes visible upon breakdown**, meaning that infrastructures are invisible until they break down (e.g. not working for maintenance reasons, server is down…)
- Infrastructure is **fixed in modular increments, not all at once or globally** for two reasons. First, because infrastructure is complex. Second, because it means different things locally, and is never changed from above.

For the scope of this white paper, however, we focus on the interplay between (i) hardware, software and data, (ii) legal & financial regulations, (iii) institutions and (iv) people, who are both using and building infrastructures.

3. Understanding trust in research practices, results and the trustworthiness of RIs

Against this background, a survey, an interview series as well as two workshops are to be highlighted: The survey - consisting of both open and closed questions that was launched among both students and professionals of Data Science - aimed at understanding the requirements for sharing and reusing open-source code better [7]. Actual research practices were compared to ideals. In addition, topics such as indicators for quality, or rather fitness for purpose, trust in open-source code and own results as well as the willingness to accept accountability were explored. Interim findings suggested that - although an uneasiness to define quality indicators was expressed - documentation was key in facilitating both quality and trust. At the same time documentation could not be clearly defined. Rather, various topics and concepts such as “good quality”, “popularity”, or “transparency” were associated with it.

Other quality indicators include an active community, the number of downloads, or test protocols. It should be noted here, that the quality indicators as determined in the quantitative analysis (participants were asked to rank a couple of indicators and explain their choices), slightly differ from quality indicators as determined in the qualitative analysis. Documentation, however, was on top in both rankings.

Last, but not least, a discrepancy between actual research practices and an ideal was discovered. Many participants stated that reusers are responsible for testing open-source code before reusing it in their own research because they - as the ones sharing it - could not be aware of future purposes. At the same time, however, many failed to live up to that re-using open-source code without running any tests or quality checks. Consequently, the willingness to accept accountability was rather low.

The semi-structured interview series was conducted by the EOSC Support Office Austria Working Group Researcher Engagement in Austria. In total, 12 researchers located in public universities in Austria were interviewed. Scientific disciplines covered included Mathematics, Computer Sciences, Life Sciences, Medical Research, Social Sciences and Humanities.

The interviews were based on a guideline focusing on two main themes, namely actual research practices from data collection to data pre-processing to analysis and interpretation as well as trust in data quality with a focus on data sharing and reuse. Findings support some of the survey’s results such as the discrepancy between actual research practices and the ideal of running test and quality checks before reusing data, services and tools as well as documentation being key for facilitating both trust and quality. Additionally, they suggest that reputation is essential for people, institutions, and infrastructures for being perceived as trustworthy. Results will be presented in detail at the STS Conference Graz 2025 [8].

In jointly organized workshops by ERCIM and the Japan Science and Technology Agency (JST) elements and facets influencing trustworthiness were discussed [5]: In research processes, Communities of Practice (CoPs), (Human) Stakeholders, Organisations and (Technical) Infrastructures were seen as key elements in (simplified) research processes. For each of these, an extensive set of indicators for trustworthiness and quality (or rather “fitness

for purpose" in specific settings) could be identified implying that indicators for neither quality, nor trustworthiness can be absolute, but depend on context and purpose.

Examples of such indicators were identified for data and code. For example, quality indicators for data relate to provenance (including ethical correctness), correctness, completeness, FAIRness, or anonymity and pseudonymity levels, while quality indicators for code range from test cases, automatic testing and test documentation to security audits to maintainability. In addition to such indicators for data and code, research communities have begun to formalize trust-enhancing practices. For example, top-tier computer science conferences such as NeurIPS and AAAI have introduced mandatory checklists during the paper submission process, covering issues such as data provenance, code availability, reproducibility of results, and societal impacts. These checklists, along with supplementary material requirements, serve as operationalized trustworthiness indicators, incentivizing researchers to provide metadata, documentation, and verification artifacts.

The follow-up workshop focused on reviewing the initial findings to explore the influence of cultural differences. The focus was to pinpoint key research areas and issues that need to be addressed to sustain and enhance trust in data-driven research. Based on previous results, the focus shifted to indicators of trustworthiness. Indicators of trustworthiness are either tangible or observable traits such as long-term financial stability, the prominence of a tool or publication, the extent of code testing, or the number of reuse and deployment cases. While these indicators can be quantified and aggregated into a potential "trustworthiness score," actual trust is shaped by both rational evaluation and intuitive, often subconscious, interpretation. External "trust donors (auditing bodies or certifying institutions) also influence trust by validating or challenging specific indicators. Their own credibility can enhance or diminish the perceived trustworthiness of the entities they assess.

Building on the initial structuring of elements and properties of the research process we need to identify the contributing factors supporting the emergence of trust.
We focus on **entities** such as a specific institution, an infrastructure, data, software, or experimental results. These entities are awarded with a certain amount of trust, which determines their perception and hence, ultimately, fitness to be used for a specific purpose (i.e. the key definition of quality).
Each entity has associated **properties**, which are characteristics that are essential to the entity and for its use or role in a specific research process. These properties can be legal status, funding of an organization, volume of data, functionality of code, etc.
These properties are, in turn, linked to **trustworthiness indicators**, which are characteristics that can be determined for a property such as long-term financial stability; prominence of a tool, paper or institution; tests performed on code; number of code reuse reports / deployment / take-up; In a mechanistic world-view these can be measured and aggregated, allowing potentially the computation of some aggregate "**trustworthiness score**". However, the trust attributed by a stakeholder onto the properties and hence ultimately on a specific entity, will depend on the rational but also the intuitive perception and interpretation of the trustworthiness indicators and a subconscious aggregation of these.
A further role is played by external "**trust donors**", i.e. individuals or institutions that may provide attribution on trustworthiness indicators such as auditing and certification institutions,

that via their trustworthiness provide an accordingly increasing or decreasing factor to the measurements of the indicators.

This leads to the identification of three major gaps, which are

1) a lack of understanding of trustworthiness indicators and their associated metrics. While several of these concepts have been developed over the years (ranging from high-level ISO standards such as the ISO9000 or ISO27000 families of standards via specific standards in the area of research data and their management (Core Trust Seal [L1], FAIR Metrics [L2] capturing fulfillment of the FAIR principles, initiatives towards more comprehensive research output impact assessment) there is a lack of understanding in how far which of these capture which essential characteristics of the trust indicators. Which of the numerous properties / characteristics of an entity may serve in which form as a potential indicator of the degree of trustworthiness?
2) A lack of understanding how a score for the overall trustworthiness could be “computed” based on whichever measurements of the individual indicators. While numerous models for aggregating diverse measurements have been devised, we lack an understanding of how these are perceived and “weighted” by individuals, most likely heavily depending on the specific context, and equally likely significantly influenced by cultural and domain-specific biases. How could a model for such an aggregation look like that would allow machines to help researchers in selecting trustworthy sources and entities for their research in increasingly global and cross-disciplinary research processes where common sense and personal knowledge are no longer sufficient to serve this task? "What might a model for such aggregation look like - one that enables machines to assist researchers in identifying trustworthy sources and entities, especially in the context of increasingly global and cross-disciplinary research, where common sense and personal knowledge alone are no longer adequate?"
3) Last, but not least, in spite of all attempts to make quality (i.e. fitness for purpose, one of the key aspects of trustworthiness) objectively compute-able or trace-able, the awarding of trust is ultimately a social process. There is a wealth of knowledge and expertise in philosophy of how trust emerges or why it fails to emerge. Yet, how to turn this into pro-active support in the specification, creation and operation of research infrastructures, code, data or results, i.e. its influence on the design of the entities and their properties, is not sufficiently well understood. How should we design and document an experiment, a data collection, etc., which information to capture, which levels of transparency to provide in which form so that people can ultimately feel they can trust the institution, the result, or any other entity to form part of their research process or decision making?

4. Need for Action: Trust-related research to accompany RI design, development and operation

4.1 Summary of Key Points

In conclusion, three key areas warrant attention: the development of infrastructures, trust, and the implementation of effective mechanisms for self-assessing trustworthiness:

Infrastructures are playing an increasingly important role in Research because of the increasing interdisciplinarity and complexity of modern Research. Against this background, researchers are facing significant challenges in fully comprehending the methods and intricate processing pipelines underpinning their work. This fast-paced environment, characterized by the reuse and repurposing of real-time data and components, further complicates the understanding of suitability, limitations, and potential quality issues, especially when applied beyond their original scope. Consequently, the potential for accumulating errors grows, making it difficult for researchers to maintain trust in their own outputs and accept accountability for results, ultimately threatening the social acceptance of scientific findings when errors are uncovered. In this context, establishing trust against intentional tampering (e.g. security) is essential.

Additionally, in situations where comprehensive verification is impractical, **trust** becomes a crucial element influencing researchers' decisions regarding the adoption and reuse of existing scientific outputs. It's important to distinguish trustworthiness, an inherent quality, from trust, which is the belief in that quality. A significant gap exists between the ideal of rigorous testing and quality checks before reuse and the reality of limited additional scrutiny in practice, often leading to a reliance on superficial indicators like download numbers rather than genuine confidence in research results. While thorough documentation, detailing provenance, testing procedures, and evidence of quality, could foster trust and ensure fitness for purpose, it is often lacking in both quantity and quality. Furthermore, conventional indicators of trustworthiness, such as the reputation of researchers, institutions, or publication venues, are becoming less reliable in the face of an expanding global research landscape and rapid advancements. Traditional metrics like download counts, citations [9] or “likes” (social media reputation) [1] are susceptible to manipulation, further eroding their trustworthiness. Similarly, brittleness towards small perturbations in inputs harms trustworthiness [4]. Although research ethics and practices are often outlined in policy documents, there's no guarantee of their consistent implementation. In the context of AI, explainability offers insights into a system's workings but doesn't provide the same level of verifiable truth [15] as the full auditability found in safety-critical systems [3]. These factors collectively contribute to a weakening of established mechanisms for building and maintaining trust in science.

To enable the **self-assessment of trustworthiness** and to support the development of trust in AI, three key approaches can be identified: Enabling the self-assessment of trustworthiness in AI systems requires a multi-faceted approach. First, systems must provide effective and timely responses to user needs. Whether assistance is delivered by a human or an AI agent, the ability to receive accurate and relevant answers quickly plays a critical role in fostering trust. Users are more likely to place confidence in systems that demonstrate practical utility and responsiveness in addressing their queries or problems.

Second, traceability and transparency must be embedded into system design. It should be possible to easily reconstruct and understand what occurred during any interaction with the system. This applies not only to AI but also to rule-based or procedural infrastructures, which can also function as opaque black boxes. For example, failures in pipeline computations must be explainable, and the current status of research objects should be readily accessible detailing who has access, what changes were made, by whom, and when.

Finally, proactive and contextual notifications contribute significantly to trust. Systems should automatically provide users with timely updates, such as email alerts regarding system actions, upcoming issues, or significant changes in status. These notifications support transparency, promote accountability, and help ensure that users remain informed and engaged with the systems they rely on.

4.2 Recommendations

Based on these considerations, the following recommendations are proposed: A deep understanding of trust is fundamental to the effective development, deployment, and adoption of research infrastructures. Scientific accompanying research plays a critical role in shaping infrastructures that are not only technically sound but also perceived as trustworthy. Such infrastructures must support confidence in the reuse of data and tools, facilitate responsible sharing of research outputs, and enable accountability in research practices. Ultimately, fostering trust at all levels ensures that both the research community and the broader public can rely on the integrity and credibility of research outcomes.

In addition, continuous studies accompanying the evolution of infrastructures are necessary to identify and examine key topics in greater depth. Specifically, the aim is to identify discrepancies between actual and research practices with a focus on why such gaps exist, the importance and acceptance of accountability, (the often overlooked issue of) distrust, infrastructures that failed (as they do offer an opportunity for failure analysis and lessons learned) as well as indicators to facilitate trustworthiness.

With regard to the latter, it is essential to closely examine which indicators can effectively support its assessment. Relevant indicators may include the availability of test documentation and test cases for open-source software, provenance information and data quality tests applied to datasets, and comprehensive metadata on data management practices, including long-term availability guarantees, citability, and versioning to preserve outdated versions for reproducibility. Additional examples include documentation of AI model training processes (detailing the qualifications of involved personnel and the quality assurance applied to training data) as well as verified identities and credentials of actors participating in research workflows. Institutional factors also play a role, such as the sustainability and resilience of hosting organisations, which may be demonstrated through accreditation, financial stability, adherence to ethical frameworks, security audits (e.g. ISO 27000, NIST-2), and legal mandates. For data, resilience may be indicated by backup strategies and geographic mirroring to mitigate risks such as funding withdrawals leading to data center closures. Furthermore, indicators may encompass quality-of-service commitments and historical monitoring results. It is equally important to assess the extent to which these indicators can be collected automatically or must be gathered manually, how they vary across object types (e.g. code, data, processes, results), disciplines, seniority levels, and cultural or regional contexts, and how resistant they are to manipulation.

Acknowledgements

Gratitude is extended to the European Research Council for Informatics and Mathematics (ERCIM), the Japan Science and Technology Agency (JST) for bringing together the experts to elaborate the above challenges. Additional thanks go to ASEA-UNINET, EOSC Focus, and the EOSC Support Office Austria Working Group Researcher Engagement for the foundational work that informed the findings of this white paper. Their work and support provided critical insights and served as a key reference throughout the development of this analysis.

Please contact: Andreas Rauber, Technische Universität Wien & SBA-Research,
rauber@ifs.tuwien.ac.at

Links

[L1] https://www.coretrustseal.org/
[L2] https://github.com/FAIRMetrics/Metrics

References

[1] Aggarwal, A. (2016): Detecting and mitigating the effect of manipulated reputation on online social networks. In *Proceedings of the 25th international conference companion on World Wide Web:* pp. 293-297.

[2] Armstrong, T. G., Moffat, A., Webber, W. and Zobel, J. (2009): Improvements that don't add up: ad-hoc retrieval results since 1998, in *CIKM '09: Conference on Information and Knowledge Management, Hong Kong China*: pp. 601-610. https://doi.org/10.1145/1645953.1646031

[3] Baniecki, H. and Biecek, P. (2024): Adversarial attacks and defenses in explainable artificial intelligence: A survey. *Information Fusion*, 102303.

[4] Biggio, B. and Roli, F. (2018): Wild patterns: Ten years after the rise of adversarial machine learning. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* : pp. 2154-2156.

[5] Emura, K. and Plexousakis, D. (2024): Report on the 4th Joint JST/ERCIM Workshop, in *ERCIM News* 136: pp. 6-10. https://ercim-news.ercim.eu/en136/jea/report-on-the-4th-joint-jst-ercim-workshop

[6] Ferrari Dacrema, M., Boglio, S., Cremonesi, P. and Jannach, D. (2021): A Troubling Analysis of Reproducibility and Progress in Recommender Systems Research, in *ACM Transactions on Information Systems*, 39(2): pp. 1-49. https://doi.org/10.1145/3434185

[7] Flicker, K., Rauber, A., Ekaputra, F.J. and Kern, T. (2024): Factors influencing Perceptions of Trust in Data Infrastructures, in *International Journal of Digital Curation*, Vol. 18(1). https://doi.org/10.2218/ijdc.v18i1.921

[8] Flicker, K., Rauber, A., Blumesberger, S., Czuray, M., Reichmann, S., Rey Mazon, M. and Saurugger, B. (2025): Identifying mechanisms to support trust in research practices & the trustworthiness of infrastructures. STS Conference, Graz 2025, Austria.

[9] Fong, E. A., and Wilhite, A. W. (2017): Authorship and citation manipulation in academic research. *PloS one*, *12*(12), e0187394.

[10] Harvey, L. and Green, D., (1993): Defining quality, *Assessment and Evaluation in Higher Education*, 18(1). pp. 9–34.

[11] Hawley, K. (2014): Trust, Distrust, and Commitment, in Sosa, E. (Eds.), *Nous* 48(1): pp. 1-20. https://doi.org/10.1111/nous.12000

[12] McLeod, C. (2020): Trust, in Edward N. Zalta (Eds.), *The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab*, Stanford University.

[13] Slota, S. C. and Bowker, G. C. (2017): How Infrastructures Matter, in Felt, U., Fouché, R., Miller, C. A., Smith-Doerr, L. (Eds.), *The Handbook of Science and Technology Studies*, MIT Press: pp. 529-554.

[14] Star, S. L. (1999): The Ethnography of Infrastructure, in *American Behavioral Scientist*, 43(9): pp. 377-391. https://doi.org/10.1177/00027649921955326

[15] Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., and Zhu, J. (2019): Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II 8*. Springer International Publishing: pp. 563-574.