Measurement Fidelity in AI-Assisted Experimental Measurement Systems.

Dea Nur Hafifah[1*], *Arita* Marini[1], *Yinghuei* Chen[2], and *Bramianto* Setiawan[3]

[1]Universitas Negeri Jakarta, 13220, Jakarta, Indonesia, [2]Asia University, Taichung, 41354, Taiwan
[3]Universitas Pelita Bangsa, Bekasi, 17530, Indonesia

Abstract. Artificial intelligence (AI) is increasingly integrated into experimental physics for detector reconstruction, signal interpretation, and automated measurement processing. However, how AI-assisted systems preserve measurement fidelity during automated experimental interpretation remains insufficiently understood. This study systematically reviewed measurement fidelity in AI-assisted physics experimentation using the PRISMA 2020 framework. Literature published between 2020 and 2025 was retrieved from Scopus, Web of Science, and IEEE Xplore, resulting in thirteen studies for qualitative synthesis. The analysis identified that measurement fidelity was primarily evaluated through detector benchmarking, uncertainty calibration, simulation-to-experiment comparison, and validation against independent reference measurements. Recurring challenges included simulation dependence, incomplete physical representation, detector-boundary effects, and uncertainty instability, indicating that computational consistency does not necessarily ensure experimentally reliable interpretation. An integrated conceptual framework is proposed to connect fidelity dimensions, validation approaches, and validation limitations, providing a conceptual basis for more reliable validation strategies in AI-assisted physics experimentation.

1 Introduction

Artificial intelligence (AI) has become increasingly integrated into experimental physics, supporting detector reconstruction, signal interpretation, simulation-assisted analysis, and automated measurement processing [1]. Modern experimental physics relies on detector instrumentation and computational measurement systems to acquire, reconstruct, and interpret complex physical signals. As AI becomes embedded within these measurement workflows, ensuring experimentally validated detector response and measurement consistency has become increasingly important for reliable physical interpretation. These advances have substantially improved the efficiency of experimental data analysis while enabling increasingly automated measurement workflows. However, computational performance alone does not guarantee physically reliable interpretation, highlighting the growing importance of experimentally grounded validation.

* Corresponding author: dea_1113825010@mhs.unj.ac.id

From the perspective of measurement science, reliable experimental interpretation depends on validated measurement models and appropriately evaluated measurement uncertainty [2,3]. These principles establish the foundation for experimental validation by ensuring that measurement results remain physically meaningful, reproducible, and comparable under defined experimental conditions. In this study, measurement fidelity is conceptualized as the ability of AI-assisted experimental measurement systems to preserve physically consistent and independently verifiable measurement interpretation throughout automated workflows. Accordingly, measurement fidelity extends beyond predictive performance by emphasizing experimentally reliable physical interpretation supported by rigorous validation procedures.

Recent advances in AI-assisted detector and measurement systems have improved detector reconstruction, simulation-assisted analysis, and uncertainty-aware modelling [1,3,4]. Nevertheless, most existing studies primarily evaluate algorithmic performance or individual computational capabilities rather than how reliable physical interpretation is maintained throughout the experimental measurement process. Consequently, comparatively little attention has been devoted to measurement fidelity as an integrated experimental validation construct that connects fidelity dimensions, validation approaches, and validation limitations. This fragmentation limits a comprehensive understanding of how experimentally reliable measurement interpretation can be preserved across increasingly automated measurement systems.

This study addresses this gap by synthesizing recent evidence on measurement fidelity in AI-assisted experimental physics. Specifically, it examines (i) the fidelity dimensions used to characterize reliable measurement interpretation, (ii) the validation approaches employed to evaluate these dimensions, and (iii) the remaining validation limitations reported in contemporary AI-assisted experimental measurement systems. The review addresses the following research questions:

1. How is measurement fidelity evaluated and validated within AI-assisted experimental physics systems?
2. What unresolved fidelity-preservation challenges emerge when experimental interpretation increasingly depends on automated measurement workflows?

Using the PRISMA 2020 framework, studies published between 2020 and 2025 were systematically synthesized to develop an integrated conceptual understanding of the relationships among fidelity dimensions, validation approaches, validation limitations, and reliable physical interpretation.

2 Method

This study employed a systematic review following the PRISMA 2020 guidelines to investigate how measurement fidelity is evaluated and preserved in AI-assisted experimental measurement systems. The review focused exclusively on peer-reviewed studies involving detector instrumentation, computational measurement systems, or AI-assisted experimental physics with explicit experimental validation or measurement-fidelity procedures.

A systematic literature search was conducted in Scopus, Web of Science, and IEEE Xplore to identify studies published between 2020 and 2025. These databases were selected because of their broad coverage of experimental physics, detector instrumentation, computational physics, and AI-assisted measurement research. To ensure that fidelity-oriented studies were explicitly identified, the search strategy combined three groups of keywords related to artificial intelligence, experimental physics, and measurement fidelity, including terms associated with experimental validation, measurement uncertainty, and physical consistency. The search query was formulated as follows:

("artificial intelligence" OR "AI-assisted" OR "AI-based" OR "machine learning") AND ("experimental physics" OR detector OR instrumentation OR "experimental measurement" OR "measurement system") AND ("measurement fidelity" OR fidelity OR "measurement validation" OR "experimental validation" OR "measurement uncertainty" OR reliability OR "physical consistency")

The search identified 274 records. After removing 37 duplicates, 237 records remained. Following exclusion based on document type, publication status, and language, 136 studies underwent title and abstract screening. Twenty-three reports were retrieved for full-text assessment, of which seven were excluded because the full text was unavailable or lacked sufficient experimental information. The remaining sixteen studies were assessed for eligibility, and three were excluded because they did not explicitly report experimentally grounded measurement validation or measurement-fidelity procedures. Consequently, thirteen studies were included in the final qualitative synthesis.

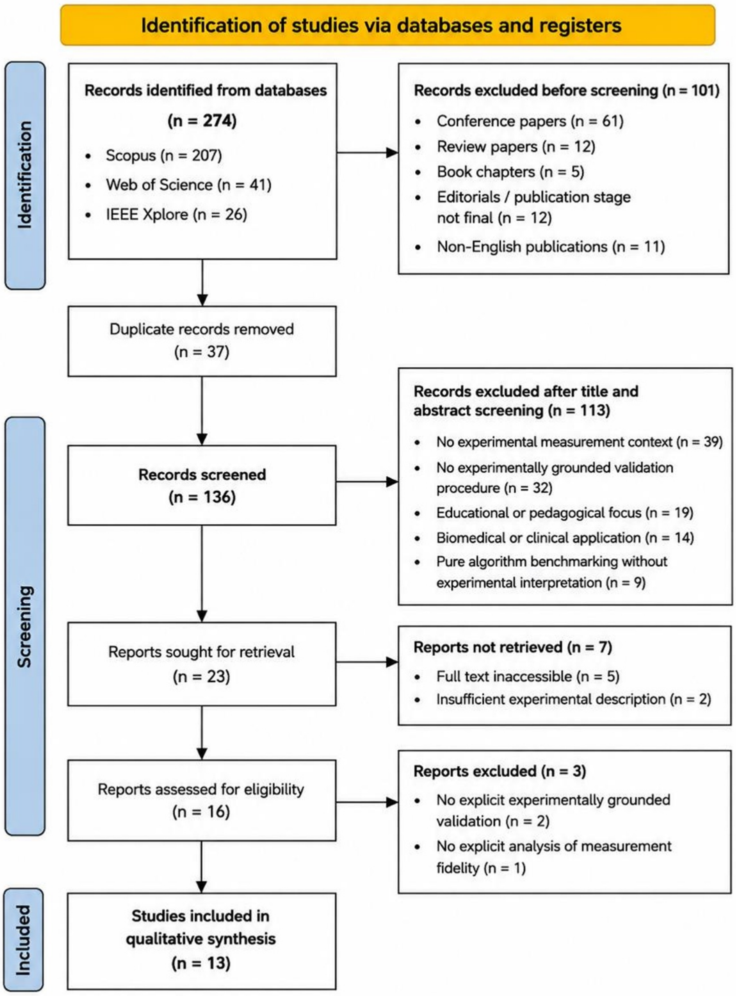

Fig 1. PRISMA flow diagram of study selection

To ensure methodological rigor, the included studies were appraised using review-specific quality criteria derived from principles of measurement science and experimental validation. The appraisal evaluated whether each study: (i) clearly defined the experimental measurement context; (ii) implemented AI-assisted detector reconstruction, measurement, or experimental interpretation; (iii) transparently described experimentally grounded validation procedures; and (iv) provided evidence supporting measurement fidelity, measurement uncertainty, or physical consistency. Only studies satisfying all appraisal criteria were included in the qualitative synthesis.

Data were extracted according to four analytical dimensions: measurement context, fidelity dimension, validation method, and key limitation. A narrative synthesis was then conducted to identify recurring relationships among fidelity dimensions, validation approaches, and validation limitations. Rather than comparing algorithmic performance alone, the synthesis examined how AI-assisted experimental measurement systems preserve physically reliable interpretation through experimentally grounded validation procedures. Particular attention was given to detector benchmarking, uncertainty calibration, simulation-to-experiment consistency, reproducibility assessment, and independent reference measurements as complementary approaches for evaluating measurement fidelity.

3 Result and Discussion

3.1 Measurement Fidelity Across AI-Assisted Experimental Measurement Systems

The thirteen included studies covered diverse experimental domains, including detector reconstruction, beam instrumentation, radiation measurement, simulation-assisted modelling, virtual experimentation, and uncertainty-aware inference. To facilitate comparison, the studies were synthesized according to four analytical dimensions: measurement context, fidelity dimension, validation method, and key limitation.

Table 1. Summary of measurement fidelity characteristics identified across AI-assisted experimental measurement studies.

Study	Measurement Context	Fidelity Dimension	Validation Method	Key Limitation
Tyson et al.[5]	Particle-detector triggering	Reconstruction fidelity	Detector-event consistency assessment	Validation emphasized classification efficiency rather than physically constrained event behaviour
Zhao et al. [6]	Drift-chamber waveform reconstruction	Signal fidelity	Experimental waveform cross-comparison	Persistent mismatch between simulated and experimentally observed signals
Ghimire et al. [7]	Nuclear scattering measurement	Background-signal fidelity	Contaminant-signal separation benchmarking	Background fluctuations reduced reliability of low-intensity signal interpretation
Tikkanen et al. [8]	Beam-profile measurement	Detector fidelity	Monte Carlo-corrected detector comparison	Residual systematic deviation remained after correction
Arpaia et al. [1]	CERN beam diagnostics	Beam-measurement fidelity	Beam-measurement cross-validation	Latent BPM faults remained difficult to isolate during dynamic operation
Beever et al. [4]	Liquid-argon detector simulation	Simulation fidelity	Experimental swarm-parameter benchmarking	Incomplete representation of non-linear recombination behaviour

Study	Measurement Context	Fidelity Dimension	Validation Method	Key Limitation
Psaros et al. [9]	Physics-informed uncertainty modelling	Uncertainty fidelity	Probabilistic uncertainty calibration	Calibration robustness deteriorated under model misspecification
Chen et al. [10]	Virtual thermoplastic experimentation	Material-response fidelity	Stress–strain response validation	Mechanical-response consistency weakened across multi-scale regimes
Minor et al. [11]	Experimental video microscopy	Spatiotemporal fidelity	Simulation-to-experiment comparison	Limited reconstruction generalization under divergent physical dynamics
Eller et al. [12]	Neutrino and water-Cherenkov event reconstruction	Reconstruction fidelity	Likelihood-free surrogate modelling	Reduced representation of fine-grained spatial correlations
Daniel et al. [13]	Gamma-photon interaction reconstruction	Localization fidelity	Physics-informed neural-network reconstruction	Accuracy deteriorated near detector boundaries
Benjamin et al. [14]	High-energy physics detector parametrization	Detector-response fidelity	Geant4-trained deep-learning reconstruction	Reliability depended on training phase-space coverage
Huang et al.[15]	Neutrino unfolding and detector correction	Unfolding fidelity	OmniFold iterative reweighting	Reweighting instability under large simulation-experiment discrepancies

The qualitative synthesis identified multiple dimensions of measurement fidelity spanning detector, signal, reconstruction, simulation, uncertainty, localization, and measurement processes. Although these dimensions emerged from different experimental contexts, they consistently reflected the ability of AI-assisted systems to preserve physically reliable interpretation rather than computational performance alone.

Across the reviewed studies, fidelity was primarily evaluated through experimentally grounded validation procedures, including detector benchmarking, waveform comparison, Monte Carlo-corrected validation, uncertainty calibration, and independent reference measurements [1,4,8]. Several studies further incorporated probabilistic calibration and physics-informed inference to strengthen robustness under partially observable conditions [9,13]. Despite these methodological differences, fidelity evaluation remained largely localized within individual experimental tasks, with limited integration across complementary validation procedures. Consequently, measurement fidelity was commonly inferred from agreement between reconstructed outputs and experimental observations rather than comprehensive experimental verification.

3.2 Fidelity-Preservation Limitations in Automated Interpretation

Despite advances in AI-assisted experimentation, three recurring challenges consistently limited measurement fidelity: simulation dependence, incomplete physical representation, and uncertainty robustness. Reconstruction reliability frequently deteriorated when experimental conditions deviated from simulation models or training data [4,11,15]. Several studies also reported incomplete representation of detector physics and unstable uncertainty estimation, limiting physically reliable interpretation despite stable computational performance [9,10,13].

These findings indicate that the principal challenge is no longer computational capability alone but the preservation of experimentally verifiable physical interpretation throughout increasingly automated measurement workflows. Collectively, these recurring limitations motivate the integrated conceptual framework presented in Figure 2.

3.3 Integrated Conceptual Framework of Measurement Fidelity

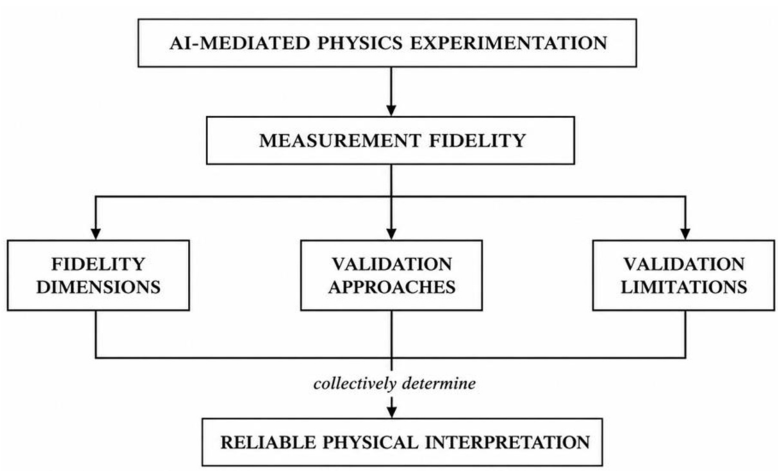

Figure 2. Conceptual framework illustrating the relationships among fidelity dimensions, validation approaches, validation limitations, and reliable physical interpretation in AI-assisted experimental measurement systems.

Figure 2 integrates the evidence synthesized from the reviewed studies into a unified conceptual framework. Measurement fidelity is positioned as the central construct linking fidelity dimensions, validation approaches, and validation limitations, which collectively determine reliable physical interpretation. Rather than relying solely on computational performance, the framework emphasizes experimentally grounded validation as the foundation for trustworthy AI-assisted experimental measurement systems.

4 Conclusion

Measurement fidelity in AI-assisted experimental measurement systems remains predominantly evaluated through localized validation procedures, including detector benchmarking, uncertainty calibration, simulation-to-experiment comparison, and independent reference measurements. The reviewed evidence indicates that computational performance alone is insufficient to ensure physically reliable interpretation, particularly under simulation dependence and uncertainty-related limitations. The proposed conceptual framework integrates fidelity dimensions, validation approaches, and validation limitations, providing a foundation for more experimentally grounded validation strategies in AI-assisted experimental measurement systems.

This publication was funded by the Indonesia Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology, and managed under the EQUITY Program (Contract Number 4308/B3/DT.03.08/2025 and B/284/UN39/HK.07.00/2025).

References

1. P. Arpaia, G. Azzopardi, F. Blanc, et al., IEEE Instrum. Meas. Mag. 24, 47 (2021). https://doi.org/10.1109/MIM.2021.9620023
2. J. Tikkanen, S. Kirschenmann, N. Kramarenko, et al., Radiat. Meas. 165, 106949 (2023). https://doi.org/10.1016/j.radmeas.2023.106949
3. Z. Beever, D. Caratelli, A. Fava, F. Pietropaolo, F. Stocker, J. Zettlemoyer, Comput. Phys. Commun. 297, 109056 (2024). https://doi.org/10.1016/j.cpc.2023.109056
4. G. Zhao, L. Wu, F. Grancagnolo, et al., Comput. Phys. Commun. 300, 109208 (2024). https://doi.org/10.1016/j.cpc.2024.109208
5. E. Begoli, T. Bhattacharya, D. Kusnezov, Nat. Mach. Intell. 1, 20 (2019). https://doi.org/10.1038/s42256-018-0004-1
6. C. Rudin, Nat. Mach. Intell. 1, 206 (2019). https://doi.org/10.1038/s42256-019-0048-x
7. R. Tyson, G. Gavalian, D.G. Ireland, B. McKinnon, Comput. Phys. Commun. 290, 108783 (2023). https://doi.org/10.1016/j.cpc.2023.108783
8. R. Ghimire, A. Ratkiewicz, S.D. Pain, et al., Nucl. Instrum. Methods Phys. Res. B 561, 165649 (2025). https://doi.org/10.1016/j.nimb.2025.165649
9. A.F. Psaros, X. Meng, Z. Zou, L. Guo, G.E. Karniadakis, J. Comput. Phys. 477, 111902 (2023). https://doi.org/10.1016/j.jcp.2022.111902
10. Z. Chen, Y. Wu, Y. Xie, K. Sattari, J. Lin, Mater. Horiz. 11, 6028 (2024). https://doi.org/10.1039/D4MH01022A
11. E.N. Minor, S.D. Howard, A.A.S. Green, et al., Soft Matter 16, 1751 (2020). https://doi.org/10.1039/C9SM01979K
12. P. Eller, A.T. Fienberg, J. Weldert, et al., Nucl. Instrum. Methods Phys. Res. A 1048, 168011 (2023). https://doi.org/10.1016/j.nima.2023.168011
13. G. Daniel, M.B. Yahiaoui, C. Comtat, et al., Eng. Appl. Artif. Intell. 131, 107876 (2024). https://doi.org/10.1016/j.engappai.2024.107876
14. D. Benjamin, S. Chekanov, W. Hopkins, Y. Li, J.R. Love, J. Instrum. 15, P05025 (2020). https://doi.org/10.1088/1748-0221/15/05/P05025
15. R.G. Huang, A. Cudd, M. Kawaue, et al., Phys. Rev. D 112, 012008 (2025). https://doi.org/10.1103/sp1f-n9k2