International Journal of Engineering and Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 15 Issue 7 July 2026 || PP 18-26

Robot Execution Assurance in Self-Driving Laboratories

Pablo Salazar-Villacis
School of Aeronautical, Automotive, Chemical and Materials Engineering (AACME), Loughborough University, Loughborough, UK

ABSTRACT: *Self-driving laboratories (SDLs) integrate robotic manipulation, analytical instrumentation and machine learning to accelerate research in chemistry and materials science. Scientific validity nevertheless depends on whether recorded execution states represent the physical operations underlying subsequent analyses. Command completion, provenance records and plausible instrument outputs can coexist with sample misidentification, incomplete transfer, timing deviation, contamination or history-altering recovery. Research spanning laboratory robotics, workflow orchestration, provenance, runtime monitoring, metrology and failure-aware optimisation has advanced device integration and operational traceability, yet the relationship between execution evidence and data suitability remains underdeveloped. Robot execution assurance denotes the coordinated specification, verification, monitoring and evaluation required to determine whether a robot-mediated operation can support a defined scientific use. This use-specific approach recognises that recovery may preserve endpoint composition while invalidating kinetic interpretation. Recurrent failures in identity continuity, acknowledgement, timing, monitoring and recovery establish requirements for explicit physical outcomes, verified operating conditions, validated observations and documented data-use decisions. Empirical priorities include controlled fault injection, condition-specific monitor validation, complete attempt records and interoperable evidence representations. Integrating these functions into SDL architectures would strengthen reproducibility and reduce invalid model updates in autonomous experimentation.*

KEYWORDS: *autonomous experimentation, execution assurance, laboratory robotics, runtime monitoring, self-driving laboratories*

Date of Submission: 12-07-2026 Date of acceptance: 23-07-2026

I. INTRODUCTION

Self-driving laboratories (SDLs) integrate automated experimentation, robotic manipulation, analytical instrumentation and machine learning to accelerate research in chemistry and materials science [1]–[5]. Mobile robotic chemists, automated synthesis platforms and modular closed-loop systems have demonstrated increasingly complex experimental programmes with limited continuous human intervention [6]–[11]. These developments increase experimental throughput and extend the range of conditions accessible to autonomous investigation. They also transfer responsibility for verifying and documenting physical operations from direct human observation to the laboratory's sensing, control and information infrastructure.

Workflow progression alone cannot establish that an intended experiment occurred. A liquid handler may report successful completion after aspirating air; a robot may deliver the designated container while its contents are misidentified; or a recovery action may restore a final measurement while altering reaction timing or contamination history. In each case, the workflow may continue and the instrument may return a plausible result even though the physical experiment differs from the experiment represented in the dataset. When such outputs enter an optimisation or machine-learning loop without qualification, operational failures may be interpreted as valid scientific observations and influence subsequent experimental decisions.

Existing SDL technologies address important components of this problem but do not yet provide an integrated basis for evaluating individual robot-mediated operations. Device interfaces communicate commands, states and errors; orchestration systems coordinate workflows and resources; runtime monitors detect selected deviations; safety systems constrain hazardous behaviour; and provenance systems preserve data lineage and responsibility. Laboratory quality systems and metrological practice further establish requirements for method validation, traceability, uncertainty and decision rules [12]–[14]. These capabilities answer different questions. Controller completion records a software state without necessarily establishing the required physical outcome. Provenance preserves the origin and derivation of a record without independently validating the event represented by that record. Similarly, the absence of an alarm provides meaningful evidence only within the validated coverage and operating conditions of the monitoring system.

Robot execution assurance is defined here as the specification of required physical outcomes, verification of prerequisite conditions, monitoring of physical execution and evaluation of the resulting evidence against the requirements of a defined scientific use. This definition makes reliability use-dependent rather than reducible to a single execution label. Evidence from a recovered operation may, for example, support sample identity and final composition while leaving transferred amount, contamination status or timing unresolved. The same operation may consequently remain suitable for endpoint comparison but be unsuitable for kinetic analysis. Evaluation must therefore distinguish among supported, contradicted and unresolved properties instead of treating workflow completion as sufficient evidence of scientific validity.

The analysis focuses on the interface between robot-mediated execution and scientific interpretation, particularly in material transfer, sample identity, instrument handover and fault recovery. It examines the capabilities and limitations of current SDL architectures, identifies recurrent failures in the execution–evidence chain and derives requirements for connecting physical observations and recovery histories to defensible data-use decisions. Sampling, analytical measurement, method validation and statistical modelling introduce additional sources of uncertainty, but they do not remove the need to establish whether the physical operation represented by a dataset was adequately performed and documented.

II. FROM ROBOT EXECUTION TO SCIENTIFIC EVIDENCE

Platform capability and aggregate performance do not, by themselves, establish the validity of an individual experimental result. A robotic platform may implement a given operation and exhibit high reliability under defined test conditions while the outcome of a particular execution remains uncertain. Data excluded from their intended scientific analysis may nevertheless retain value for characterising equipment performance, experimental feasibility or failure mechanisms.

Performance characterisation remains essential because laboratory materials and geometries directly affect robotic behaviour. Injection depends on alignment and tool engagement; solid dispensing on powder flow and sensing; and viscous transfer on aspiration, wetting, residue and dispensing dynamics [15]–[17]. Pouring, flexible liquid handling and automated pipetting are similarly influenced by container geometry, calibration, fluid properties and observation conditions [18]–[23]. Evidence from command completion is therefore interpretable only within experimentally validated conditions. Transferring a robotic skill to another platform or exposing it through a common interface does not demonstrate that its performance characteristics and monitoring limitations have been preserved.

For scientific evaluation, the relevant unit is the individual robot-mediated operation whose execution can affect data interpretation. Examples include material transfer, vessel sealing, sample handover, cleaning and instrument loading. This unit lies between an individual device message and the complete experiment. Repetition, delay, substitution or partial completion at this level may alter the physical history represented by the resulting data even when the surrounding workflow completes normally.

Assessment of an individual operation requires explicit correspondence among its intended physical effect, execution conditions, verifying observations and subsequent interventions. The intended effect identifies the material, destination and physical change required by the downstream analysis. Execution conditions include calibration, equipment status, resources, timing and the validated operating range. Observations provide evidence of the realised physical state, whereas the recovery record identifies retries, compensating actions and human intervention. A single success label cannot represent these dimensions: one operation may support sample identity, contradict the required timing and leave the transferred amount unresolved.

An observation supports only those physical properties that it can validly assess. A camera image may establish target position within a validated field of view without establishing chemical composition. A pressure trace may support successful aspiration without demonstrating delivery into the intended vessel. A balance may constrain mass change, but the measured value can include exterior wetting, evaporation, contact force and tare error. Similarly, an instrument acknowledgement can establish that an analytical run started without confirming presentation of the correct sample. Computer-vision systems for laboratory vessels, solubility, phase separation and extraction demonstrate the value of physical monitoring while illustrating its dependence on task, material and observation conditions [24]–[28]. Additional sensors increase confidence only when their measurements discriminate relevant failure states and their shared dependencies are recognised.

Recovery introduces an additional evidential problem because it may restore workflow progression without restoring the original experimental history. A delayed second addition may achieve the intended final amount while invalidating reaction timing. Replacement of a blocked line may restore measurement availability while introducing a new cleaning or contamination boundary. Manual relocation of a vial may preserve throughput while weakening evidence of identity or location. The scientific acceptability of a recovered result therefore depends on the properties restored by the intervention and the requirements of the intended analysis.

Scientific reliability is also distinct from safety and provenance. A safety controller can stop hazardous motion or isolate energy, but a safe state may leave the experimental outcome unresolved. A provenance system can record an operation, observation and responsible person without establishing whether the observation supports the required physical outcome. Metrological practice provides a useful analogue by separating method validation, measurement, uncertainty, specification and the decision rule used to determine conformity

[29]–[33]. Robot-mediated experimentation requires a corresponding separation among recorded events, interpretation of physical evidence and decisions about acceptable data use.

III. CAPABILITIES AND LIMITATIONS OF CURRENT SDL ARCHITECTURES

Current SDL architectures provide a substantial foundation for reliable execution, particularly at the levels of device interoperability and workflow coordination. SiLA 2 defines typed features, commands, status transitions and errors at the device interface, supported by implementation tooling for rapid integration [34], [35]. ChemOS and ChemOS 2.0 coordinate modular chemical workflows [36]–[38], whereas AlabOS adds reconfigurable workflows and resource scheduling for concurrent autonomous operation [39]. IvoryOS and PyLabRobot provide interoperable interfaces for Python-based SDLs and laboratory hardware [40], [41]. Hierarchical automation, NIMS-OS and multi-agent frameworks extend orchestration across feedback loops and shared facilities [42]–[45]. Collectively, these systems reduce integration costs and expose operational states that would otherwise remain confined to device-specific software.

A complementary body of work addresses continuity between materials, workflows and data. Distributed knowledge graphs connect material and data flows across laboratories [46], [47], while property-graph approaches support automated metadata capture and reproducibility [48]. Cross-platform chemical programming languages strengthen the repeatable execution of synthesis procedures [49], [50]. W3C PROV provides a general representation of entities, activities, agents and derivation [51], [52], building on the broader scientific-workflow provenance literature [53], [54]. FAIR data principles, community standards and open reaction repositories further support reuse and cross-platform interpretation [55]–[58]. These approaches increasingly permit results to be traced to the workflows and resources from which they originated.

A third body of work concerns runtime observation and recovery. Automated HPLC anomaly detection and open chromatographic analysis support prospective monitoring of analytical quality [59], [60]. LIRA integrates localisation, inspection and reasoning for autonomous laboratory workflows [61], while multimodal behaviour trees combine perception with reactive execution [62] within a broader robotics literature on modular reactive control [63]. Research on laboratory self-maintainability extends this focus from individual tasks to sustained autonomous operation [64]. Runtime-verification studies further establish the need to specify monitored properties, event semantics and observation coverage [65], [66]. Physical verification and recovery are therefore technically feasible, although their implications for downstream scientific use remain incompletely integrated.

Device interfaces, workflow orchestrators, provenance systems, runtime monitors, metrological controls and failure-aware optimisation provide complementary inputs to robot execution assurance. Table I distinguishes the function established by each research area from the limitation that remains when its output is used to support scientific inference.

Table I. *Capabilities of current SDL technologies and their remaining limitations for scientific inference.*

Source family	Demonstrated contribution	Limitation for scientific inference
Device interfaces [34], [35]	Typed commands, states, errors and interoperable device functions	Interface completion does not establish the required physical postcondition.
Orchestration and resources [38], [39], [40]	Workflow coordination, modular interfaces, scheduling and execution records	Execution records alone do not determine which analyses may use the output.
Provenance and reproducibility [46], [48], [49], [51]	Material and data flow, metadata, derivation and repeatable procedures	Recorded assertions require evidential support and continuing validity.
Runtime monitoring and recovery [59], [61], [62], [64]	Anomaly detection, inspection, reactive control and maintenance	Recovery may leave required physical conditions unsupported or unresolved.
Laboratory quality and metrology [29], [30], [31]	Controlled methods, traceability, uncertainty and decision rules	Decision rules must be applied to individual robot-mediated operations.
Failure-aware optimisation [67], [68]	Models for unknown feasibility and experimental failure	Platform failure, monitor failure and scientific infeasibility require distinct labels.

Current SDL architectures can schedule, execute and record an operation without establishing whether the resulting data remain suitable for kinetic modelling, endpoint comparison or model updating. Data suitability depends on the physical conditions required by the analysis, the observations available for the individual operation and any changes introduced by recovery. The principal unresolved issue, and the focus of robot execution assurance, is therefore the correspondence between execution records and downstream scientific interpretation rather than command execution alone.

Use-specific execution decisions can be incorporated into existing orchestrators, event stores, electronic laboratory notebooks and knowledge graphs rather than requiring a separate middleware platform. The essential requirement is preservation of the evidence supporting acceptance, restriction or exclusion of data, together with the capacity to revise that decision if a calibration, identity mapping or other supporting record is subsequently invalidated.

IV. FAILURE MODES IN THE EXECUTION–EVIDENCE CHAIN

Failures that produce plausible measurements from an incorrect physical history present a particular risk because they may remain undetected during routine analysis. They define the conditions under which robot execution assurance must distinguish recorded workflow progression from evidence of the physical operation. Six recurrent classes are identified: loss of identity, material-dependent execution error, correlated verification failure, ambiguous acknowledgement or timing, history-altering recovery and selective exclusion of difficult experiments (Table II).

Table II. *Recurrent failures in the execution–evidence chain and their scientific consequences.*

Failure class	Typical example	Discriminating evidence	Scientific risk
Identity discontinuity	Correct motion targets the wrong vial	Independent identity checkpoint, location history and content-specific observation	Label noise, cross-contamination or false structure-property relation
Material-dependent deviation	Viscous formulation is under-delivered	Material-specific validation plus pressure, mass or endpoint evidence	Candidate-dependent bias and distorted optimisation
Verification gap or common cause	Camera and deck map share the same wrong mapping	Declared coverage, dependency analysis and an independent physical challenge	False corroboration and systematic sample misassignment
Ambiguous acknowledgement or time	Physical action may occur after communications are lost	Persistent attempt identity, causal event links and clock-uncertainty intervals	Duplicate action, incorrect event order or invalid kinetics
Non-repeatable recovery	Delayed retry reaches the final amount	Intervention history and post-recovery evidence for each required property	Endpoint may remain usable while process-history claims fail
Selective admission	Difficult candidates are repeated or removed	Complete attempt denominators, failure mechanism and candidate descriptors	Biased response surface or misleading feasibility boundary

Identity continuity requires consistent linkage among the digital sample identifier, physical container, material contents, location, operation target and resulting data file. Concordant records do not constitute independent confirmation when they derive from a common mapping; for example, a scheduler, camera annotation and deck map may all reproduce the same erroneous assignment. Where the scientific consequence warrants additional verification, the confirming observation should be independent of the shared source. This requirement is consistent with evidence that chemical reproducibility depends on explicit procedures, abstractions and complete records rather than nominal repetition alone [13], [69].

Communication acknowledgements represent distinct software events with different evidential meanings. Receipt confirms arrival of a message; acceptance confirms that syntax and local preconditions were satisfied; initiation indicates that physical execution may have begun; and controller completion indicates that a local procedure reached its configured completion condition. None of these events independently establishes the intended physical postcondition. When communication fails after execution may have started, the absence of an acknowledgement cannot distinguish non-execution from partial or complete execution. Automatic retry under these conditions may therefore introduce a duplicate physical action.

Temporal interpretation is complicated by heterogeneous device clocks, communication latencies and event-ordering guarantees. Logical sequence numbers and causal links preserve local relations, whereas wall-clock times require explicit uncertainty intervals [70]. If the plausible completion interval crosses a scientifically relevant deadline, timing remains unresolved even when the final endpoint is acceptable. Substitution of a single timestamp for this interval would introduce unsupported temporal precision.

Monitor validity is conditional on the circumstances of observation. Cameras may be occluded, sensors may drift, thresholds may be calibrated on restricted material classes, and multiple monitors may share the same calibration or mapping. The absence of an alarm constitutes supportive evidence only when the monitor was capable of detecting the relevant failure under the prevailing conditions. Studies of autonomous viscosity estimation and laboratory vision demonstrate that sensing performance depends on material and scene characteristics [24], [28], [71]. Candidate properties may therefore influence both the scientific response and the probability of successful execution or observation; dark, reflective, viscous, foaming or precipitating samples are salient examples.

Candidate-dependent execution or monitoring does not necessarily produce statistical bias. If data inclusion depends only on observed candidate variables, the response model is correctly specified and adequate support remains, unequal inclusion may primarily reduce precision. Distortion becomes more likely when inclusion depends on an unobserved determinant of response, some regions receive negligible support, the model is misspecified, repeated screening changes the effective intervention or an adaptive policy learns only from selectively retained outcomes [72], [73]. Failed experiments can contain information that is absent from success-only datasets [74], [75]. Methods for unknown feasibility and experimental failure are consequently relevant [67], [68], provided that operational systems preserve distinct and scientifically interpretable failure labels.

Figure 1 applies these distinctions to a hypothetical transfer of 100 mg from a specified source lot with a required completion time of ten seconds. Communication is lost after physical execution may have started, precluding automatic retry. A subsequent audit constrains completion to an interval of 9.6–10.4 s, which crosses the timing criterion. Gross mass constrains the transfer but does not resolve the in-vessel amount because exterior retention remains possible. A validated lot-specific assay supports identity and final composition. Endpoint analysis may therefore be retained, whereas kinetic analysis must be excluded. The example demonstrates that recovery may re-establish one scientifically relevant property without re-establishing all properties required by the original analysis.

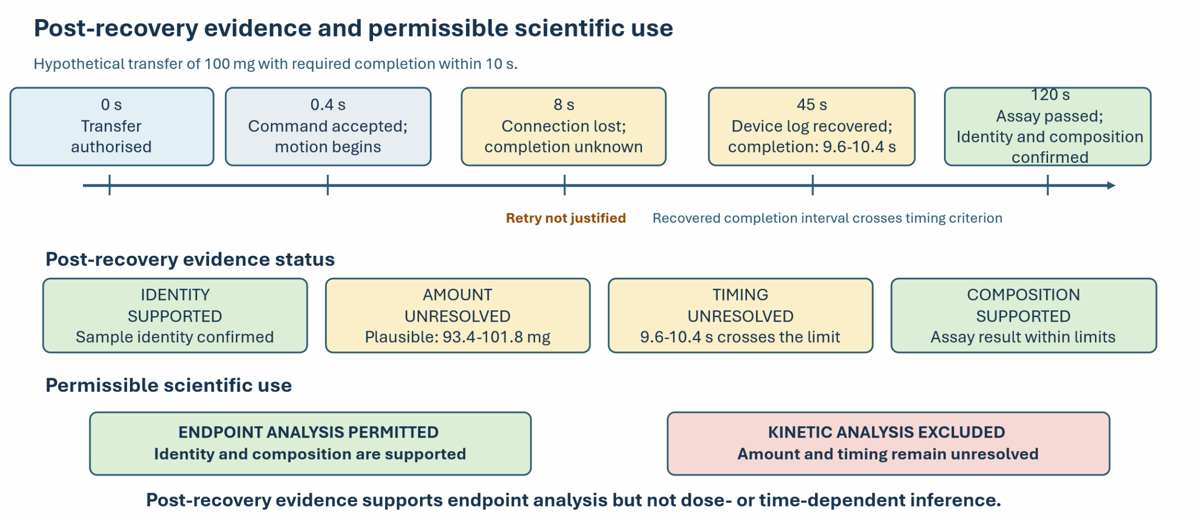

Fig. 1. Post-recovery evidence supports sample identity and final composition but not transferred amount or the 10 s timing requirement; endpoint analysis remains permissible, whereas kinetic analysis is excluded.

V. REQUIREMENTS FOR ROBOT EXECUTION ASSURANCE

Evidence across robotics, monitoring, provenance and metrology supports four interdependent functions of robot execution assurance (Fig. 2). Definition specifies the intended physical action, downstream analysis, resources, timing, required outcomes, observations and recovery limits. Verification confirms sample identity, calibration, device health, authority and resource availability before execution. Monitoring links a persistent operation identifier to controller state, physical observations, timing and deviations. Decision integrates the available evidence to determine which analyses the resulting data can support. Subsequent evidence may revise this determination without altering the original execution history.

These functions are sequentially interdependent. Verification prevents a syntactically valid command from being interpreted as authorisation to act. Monitoring prevents a controller completion flag from substituting for evidence of the required physical outcome. Determination of acceptable use prevents successful workflow recovery from implying restoration of every scientifically relevant property. Feedback to experimental planning should preserve supported, contradicted and unresolved outcomes as distinct observations.

Operational implementation requires three linked record types. The action record identifies the sample, resources, validated conditions, command, required physical outcomes, timing and intended analyses. The observation record describes the source, validity, coverage, uncertainty, shared dependencies and decision rule. The recovery record documents the trigger, risk of duplicate action, intervention, conditions for resuming work, post-recovery checks and analyses that remain possible. These records need not be stored as separate files, but their meanings should be represented by explicit and versioned fields or policies within the laboratory infrastructure.

Uncertain execution should be managed conservatively. A timeout after physical execution may have begun does not establish that no action occurred, and a non-repeatable operation with uncertain completion should not be retried automatically. Observations that share an unresolved mapping, clock or calibration cannot be treated as independent confirmation. An operation whose uncertainty crosses the boundary of validated conditions should be deferred or managed under a separately validated procedure. Recovery requires re-evaluation of every analysis that depends on an altered property, while absent observations remain unresolved rather than being interpreted favourably.

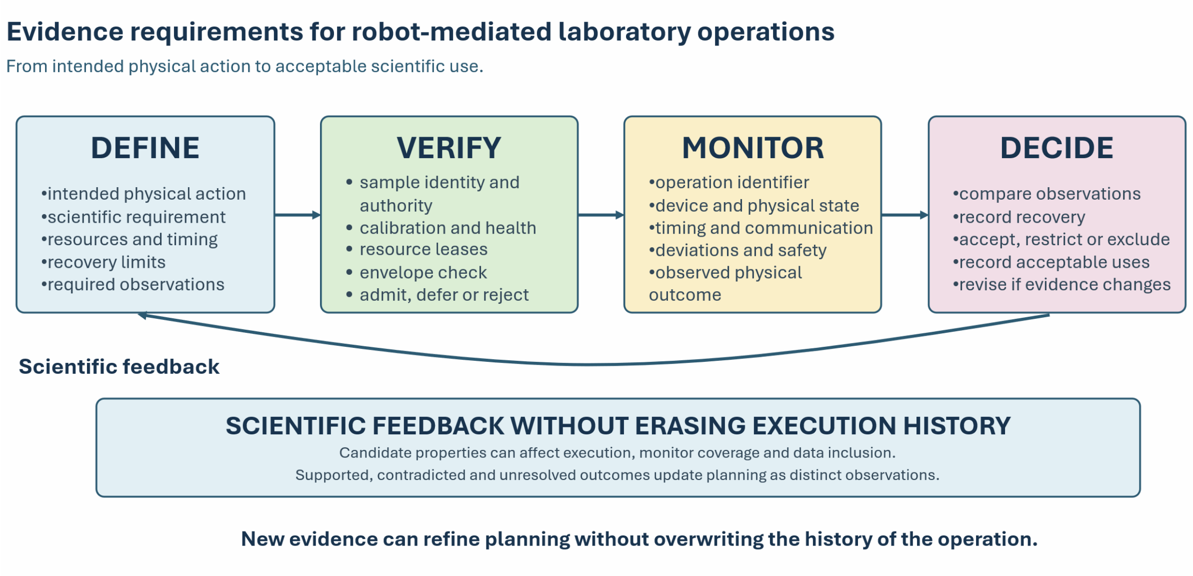

Fig. 2. Four functions of robot execution assurance connect an intended robot-mediated operation to a documented decision about acceptable data use while preserving faults, recovery and subsequent revisions.

Functional separation reduces conflicts among scientific planning, resource coordination, execution, safety and evidence evaluation. A scientific planner specifies proposed operations and intended analyses; a scheduler coordinates shared equipment; an execution service controls the operation; and a safety controller can terminate hazardous behaviour. A separate decision rule or authorised reviewer determines acceptable data use. Human intervention becomes part of the execution history. Neither safety approval nor workflow progression is sufficient to authorise scientific use of a result.

Implementation may proceed incrementally. A small SDL can initially assign persistent identifiers to a limited set of scientifically consequential operations and link them to observations, interventions and data-use decisions. Automated reaction systems, flow platforms and robotic chemistry architectures demonstrate that persistent identifiers and observations can be attached to modular execution [76]–[81]. Initial implementations need not retain every low-level robot message, but they should distinguish operations that were attempted, completed normally, completed after intervention, left unresolved or excluded from scientific analysis. Campaign-level records should also retain deviations, human interventions, exclusions and repeated operations. Candidate characteristics relevant to execution should accompany these records where scientifically and ethically appropriate.

At larger scale, these records may be distributed across an orchestrator, resource service, event store and dependency graph. Existing provenance models can represent the resulting entities and relations [51], [52], but the decision regarding acceptable data use should precede inclusion in analysis or model updating. Records must also remain revisable. If a calibration or identity mapping is subsequently invalidated, the laboratory should identify affected operations, samples, results and model versions and issue a revised decision without overwriting the earlier record.

The intensity of monitoring and documentation should be proportionate to scientific consequence. Stronger controls are warranted for non-repeatable operations, irreplaceable samples, material-dependent failure, limited monitoring coverage and recoveries that alter experimental history. Lower-burden controls may be adequate for stable, reversible and independently verified operations. The selected controls should reflect consequence, observability, reversibility, replacement cost and the potential for selection bias rather than a single maturity score.

Robot execution assurance imposes costs in latency, storage and operator workload, and unnecessary exclusion may distort a dataset when monitor performance varies across candidate classes. User interfaces and machine-readable records should therefore represent the remaining physical alternatives and the observations capable of discriminating among them. Its performance should be evaluated by the extent to which recorded evidence distinguishes supported, contradicted and unresolved outcomes while maintaining a proportionate experimental burden.

VI. OUTLOOK AND CONCLUSIONS

Empirical validation of robot execution assurance is required to determine whether monitoring and recovery preserve the physical conditions needed by downstream analyses. Controlled fault-injection studies should examine representative operations, including liquid transfer and instrument handover, under communication loss, clock disagreement, stale identity mappings, resource conflicts, partial completion, contamination, device restart, human intervention and retrospective calibration failure. Relevant outcomes include false acceptance, unnecessary exclusion, duplicate physical action, time to resolution and additional cycle time.

Monitoring performance should be reported by failure mode and operating condition rather than as an aggregate accuracy measure. Material properties, vessel geometry, illumination, equipment wear and shared calibration errors may all affect fault observability. Independent assessors should apply predefined decision rules to blinded execution records. Inter-assessor agreement would indicate that acceptable uses of experimental results can be determined reproducibly, whereas disagreement would identify ambiguous requirements or insufficient observations.

Autonomous campaigns should preserve failed, constrained, recovered and scientifically excluded attempts as distinct records. Failure-aware optimisation can use such outcomes only when scientific infeasibility is distinguished from robotic failure, monitoring failure and unresolved execution [67], [68], [75], [82]. Comparative studies are required to determine when success-only logging or candidate-dependent exclusion alters model coverage, response estimation or optimiser behaviour.

For robot execution assurance to be interoperable, common representations must extend beyond portable commands and device states. Existing interfaces communicate device functions and operational states [34], [35], but reproducibility across laboratories additionally requires consistent representations of sample identity, required physical outcomes, monitoring limitations, interventions and acceptable downstream uses. Structurally equivalent records do not constitute equivalent evidence when laboratories have different sensor coverage or apply materially different decision rules.

Controller completion and instrument output are operational records; neither independently establishes that the intended experiment was performed. Robot execution assurance addresses this gap by linking specified physical outcomes, verified prerequisites, validated observations and recovery histories to explicit decisions about acceptable scientific use. Incorporating these links into SDL architectures would enable autonomous laboratories to learn from substantiated experimental outcomes rather than from workflow completion alone.

ACKNOWLEDGEMENTS

The author acknowledges the support provided by Ecuador's former Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT), whose functions are now administered by the Ministerio de Educación, Deporte y Cultura (MINEDEC).

REFERENCES

[1] F. Häse, L. M. Roch, and A. Aspuru-Guzik, "Next-generation experimentation with self-driving laboratories," *Trends in Chemistry*, vol. 1, no. 3, pp. 282–291, 2019, doi: 10.1016/j.trechm.2019.02.007.

[2] Milad Abolhasani and Eugenia Kumacheva, "The rise of self-driving labs in chemical and materials sciences," *Nature Synthesis*, vol. 2, no. 6, pp. 483–492, 2023, doi: 10.1038/s44160-022-00231-0.

[3] G. Tom et al., "Self-driving laboratories for chemistry and materials science," *Chemical Reviews*, vol. 124, pp. 9633–9732, 2024, doi: 10.1021/acs.chemrev.4c00055.

[4] Martin Seifrid *et al.*, "Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab," *Accounts of Chemical Research*, vol. 55, no. 17, pp. 2454–2466, 2022, doi: 10.1021/acs.accounts.2c00220.

[5] Stanley Lo *et al.*, "Review of low-cost self-driving laboratories in chemistry and materials science: the 'frugal twin' concept," *Digital Discovery*, vol. 3, no. 5, pp. 842–868, 2024, doi: 10.1039/d3dd00223c.

[6] Benjamin Burger *et al.*, "A mobile robotic chemist," *Nature*, vol. 583, no. 7815, pp. 237–241, 2020, doi: 10.1038/s41586-020-2442-2.

[7] B. P. MacLeod et al., "Self-driving laboratory for accelerated discovery of thin-film materials," *Science Advances*, vol. 6, no. 20, p. eaaz8867, 2020, doi: 10.1126/sciadv.aaz8867.

[8] S. Hessam M. Mehr, Matthew Craven, Artem I. Leonov, Graham Keenan, and Leroy Cronin, "A universal system for digitization and automatic execution of the chemical synthesis literature," *Science*, vol. 370, no. 6512, pp. 101–108, 2020, doi: 10.1126/science.abc2986.

[9] N. L. Bell et al., "Autonomous execution of highly reactive chemical transformations in the Schlenkputer," *Nature Chemical Engineering*, vol. 1, pp. 180–189, 2024, doi: 10.1038/s44286-023-00024-y.

[10] T. Dai et al., "Autonomous mobile robots for exploratory synthetic chemistry," *Nature*, vol. 635, pp. 890–897, 2024, doi: 10.1038/s41586-024-08173-7.

[11] N. J. Szymanski et al., "An autonomous laboratory for the accelerated synthesis of novel materials," *Nature*, vol. 624, pp. 86–91, 2023, doi: 10.1038/s41586-023-06734-w.

[12] Melodie Christensen *et al.*, "Automation isn't automatic," *Chemical Science*, vol. 12, no. 47, pp. 15473–15490, 2021, doi: 10.1039/d1sc04588a.

[13] R. B. Canty and M. Abolhasani, "Reproducibility in automated chemistry laboratories using computer science abstractions," *Nature Synthesis*, vol. 3, pp. 1327–1339, 2024, doi: 10.1038/s44160-024-00649-8.

[14] A. A. Volk and M. Abolhasani, "Performance metrics to unleash the power of self-driving labs in chemistry and materials science," *Nature Communications*, vol. 15, p. 1378, 2024, doi: 10.1038/s41467-024-45569-5.

[15] A. Angelopoulos et al., "High-accuracy injection using a mobile manipulation robot for chemistry lab automation," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2023, pp. 10102–10109. doi: 10.1109/IROS55552.2023.10341743.

[16] Ying Jiang *et al.*, "Autonomous biomimetic solid dispensing using a dual-arm robotic manipulator," *Digital Discovery*, vol. 2, no. 6, pp. 1733–1744, 2023, doi: 10.1039/d3dd00075c.

[17] P. Quijano Velasco et al., "Optimization of liquid handling parameters for viscous liquid transfers with pipetting robots, a 'sticky situation'," *Digital Discovery*, vol. 3, pp. 1011–1020, 2024, doi: 10.1039/D3DD00255A.
[18] Monroe Kennedy, Karl Schmeckpeper, Dinesh Thakur, Chenfanfu Jiang, Vijay Kumar, and Kostas Daniilidis, "Autonomous Precision Pouring From Unknown Containers," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2317–2324, 2019, doi: 10.1109/lra.2019.2902075.
[19] Yongqiang Huang, Juan Wilches, and Yu Sun, "Robot gaining accurate pouring skills through self-supervised learning and generalization," *Robotics and Autonomous Systems*, vol. 136, p. 103692, 2021, doi: 10.1016/j.robot.2020.103692.
[20] Dennis Knobbe, Henning Zwirnmann, Moritz Eckhoff, and Sami Haddadin, "Core Processes in Intelligent Robotic Lab Assistants: Flexible Liquid Handling," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2022, pp. 2335–2342. doi: 10.1109/iros47612.2022.9981636.
[21] David C. Florian, Mateusz Odziomek, Cerie L. Ock, Hannah Chen, and Scott A. Guelcher, "Principles of computer-controlled linear motion applied to an open-source affordable liquid handler for automated micropipetting," *Scientific Reports*, vol. 10, no. 1, p. 13663, 2020, doi: 10.1038/s41598-020-70465-5.
[22] Philip Dettinger *et al.*, "Open-source personal pipetting robots with live-cell incubation and microscopy compatibility," *Nature Communications*, vol. 13, no. 1, p. 2999, 2022, doi: 10.1038/s41467-022-30643-7.
[23] Naruki Yoshikawa, Kourosh Darvish, Mohammad Ghazi Vakili, Animesh Garg, and Alán Aspuru-Guzik, "Digital pipette: open hardware for liquid transfer in self-driving laboratories," *Digital Discovery*, vol. 2, no. 6, pp. 1745–1751, 2023, doi: 10.1039/d3dd00115f.
[24] Sagi Eppel, Haoping Xu, Mor Bismuth, and Alan Aspuru-Guzik, "Computer Vision for Recognition of Materials and Vessels in Chemistry Lab Settings and the Vector-LabPics Data Set," *ACS Central Science*, vol. 6, no. 10, pp. 1743–1752, 2020, doi: 10.1021/acscentsci.0c00460.
[25] Parisa Shiri *et al.*, "Automated solubility screening platform using computer vision," *iScience*, vol. 24, no. 3, p. 102176, 2021, doi: 10.1016/j.isci.2021.102176.
[26] Rama El-khawaldeh *et al.*, "From eyes to cameras: Computer vision for high-throughput liquid-liquid separation," *Device*, vol. 2, no. 7, p. 100404, 2024, doi: 10.1016/j.device.2024.100404.
[27] Alexandra C. Sun *et al.*, "Vision-Guided Automation Platform for Liquid–Liquid Extraction and Workup Development," *Organic Process Research & Development*, vol. 27, no. 11, pp. 1954–1964, 2023, doi: 10.1021/acs.oprd.3c00217.
[28] Rama El-khawaldeh *et al.*, "Keeping an 'eye' on the experiment: computer vision for real-time monitoring and control," *Chemical Science*, vol. 15, no. 4, pp. 1271–1282, 2024, doi: 10.1039/d3sc05491h.
[29] International Organization for Standardization, *General Requirements for the Competence of Testing and Calibration Laboratories*, ISO/IEC 17025:2017, 2017. Available: https://www.iso.org/standard/66912.html
[30] Joint Committee for Guides in Metrology, "Evaluation of Measurement Data: Guide to the Expression of Uncertainty in Measurement," JCGM 100:2008, 2008. doi: 10.59161/JCGM100-2008E.
[31] Joint Committee for Guides in Metrology, "Evaluation of Measurement Data: The Role of Measurement Uncertainty in Conformity Assessment," JCGM 106:2012, 2012. doi: 10.59161/JCGM106-2012.
[32] M. Thompson, S. L. R. Ellison, and R. Wood, "Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report)," *Pure and Applied Chemistry*, vol. 74, no. 5, pp. 835–855, 2002, doi: 10.1351/pac200274050835.
[33] L. R. Pendrill, "Using measurement uncertainty in decision-making and conformity assessment," *Metrologia*, vol. 51, no. 4, pp. S206–S218, 2014, doi: 10.1088/0026-1394/51/4/S206.
[34] SiLA, "SiLA 2 Part (A): Overview, Concepts and Core Specification," Association Consortium Standardization in Lab Automation, 2022. Available: https://sila-standard.com/wp-content/uploads/2022/03/SiLA-2-Part-A-Overview-Concepts-and-Core-Specification-v1.1.pdf
[35] Lukas Bromig, David Leiter, Alexandru-Virgil Mardale, Nikolas von den Eichen, Emmeran Bieringer, and Dirk Weuster-Botz, "The SiLA 2 Manager for rapid device integration and workflow automation," *SoftwareX*, vol. 17, p. 100991, 2022, doi: 10.1016/j.softx.2022.100991.
[36] L. M. Roch et al., "ChemOS: Orchestrating autonomous experimentation," *Science Robotics*, vol. 3, no. 19, p. eaat5559, 2018, doi: 10.1126/scirobotics.aat5559.
[37] Loïc M. Roch *et al.*, "ChemOS: An orchestration software to democratize autonomous discovery," *PLOS ONE*, vol. 15, no. 4, p. e0229862, 2020, doi: 10.1371/journal.pone.0229862.
[38] Malcolm Sim *et al.*, "ChemOS 2.0: An orchestration architecture for chemical self-driving laboratories," *Matter*, vol. 7, no. 9, pp. 2959–2977, 2024, doi: 10.1016/j.matt.2024.04.022.
[39] Y. Fei et al., "AlabOS: A Python-based reconfigurable workflow management framework for autonomous laboratories," *Digital Discovery*, vol. 3, pp. 2275–2288, 2024, doi: 10.1039/D4DD00129J.
[40] W. Zhang et al., "IvoryOS: An interoperable web interface for orchestrating Python-based self-driving laboratories," *Nature Communications*, vol. 16, p. 5182, 2025, doi: 10.1038/s41467-025-60514-w.
[41] Rick P. Wierenga, Stefan M. Golas, Wilson Ho, Connor W. Coley, and Kevin M. Esvelt, "PyLabRobot: An open-source, hardware-agnostic interface for liquid-handling robots and accessories," *Device*, vol. 1, no. 4, p. 100111, 2023, doi: 10.1016/j.device.2023.100111.
[42] Fuzhan Rahmanian *et al.*, "Enabling Modular Autonomous Feedback Loops in Materials Science through Hierarchical Experimental Laboratory Automation and Orchestration," *Advanced Materials Interfaces*, vol. 9, no. 8, p. 2101987, 2022, doi: 10.1002/admi.202101987.
[43] Ryo Tamura, Koji Tsuda, and Shoichi Matsuda, "NIMS-OS: an automation software to implement a closed loop between artificial intelligence and robotic experiments in materials science," *Science and Technology of Advanced Materials: Methods*, vol. 3, no. 1, p. 2232297, 2023, doi: 10.1080/27660400.2023.2232297.
[44] A. Gilad Kusne and Austin McDannald, "Scalable multi-agent lab framework for lab optimization," *Matter*, vol. 6, no. 6, pp. 1880–1893, 2023, doi: 10.1016/j.matt.2023.03.022.
[45] Dan Guevarra *et al.*, "Orchestrating nimble experiments across interconnected labs," *Digital Discovery*, vol. 2, no. 6, pp. 1806–1812, 2023, doi: 10.1039/d3dd00166k.
[46] Jiaru Bai *et al.*, "A dynamic knowledge graph approach to distributed self-driving laboratories," *Nature Communications*, vol. 15, no. 1, p. 462, 2024, doi: 10.1038/s41467-023-44599-9.
[47] Jiaru Bai, Liwei Cao, Sebastian Mosbach, Jethro Akroyd, Alexei A. Lapkin, and Markus Kraft, "From Platform to Knowledge Graph: Evolution of Laboratory Automation," *JACS Au*, vol. 2, no. 2, pp. 292–309, 2022, doi: 10.1021/jacsau.1c00438.
[48] F. M. Mione et al., "A property graph schema for automated metadata capture, reproducibility and knowledge discovery in high-throughput bioprocess development," *Digital Discovery*, vol. 4, pp. 2401–2422, 2025, doi: 10.1039/D5DD00070J.

[49] R. Rauschen et al., “Universal chemical programming language for robotic synthesis repeatability,” *Nature Synthesis*, vol. 3, pp. 488–496, 2024, doi: 10.1038/s44160-023-00473-6.
[50] Sebastian Steiner *et al.*, “Organic synthesis in a modular robotic system driven by a chemical programming language,” *Science*, vol. 363, no. 6423, p. eaav2211, 2019, doi: 10.1126/science.aav2211.
[51] World Wide Web Consortium, “PROV-DM: The PROV Data Model,” W3C, Apr. 2013. Available: https://www.w3.org/TR/prov-dm/
[52] L. Moreau et al., “The rationale of PROV,” *Journal of Web Semantics*, vol. 35, pp. 235–257, 2015, doi: 10.1016/j.websem.2015.04.001.
[53] Y. L. Simmhan, B. Plale, and D. Gannon, “A survey of data provenance in e-science,” *ACM SIGMOD Record*, vol. 34, no. 3, pp. 31–36, 2005, doi: 10.1145/1084805.1084812.
[54] S. B. Davidson and J. Freire, “Provenance and scientific workflows: challenges and opportunities,” in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 2008, pp. 1345–1350. doi: 10.1145/1376616.1376772.
[55] Mark D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, no. 1, p. 160018, 2016, doi: 10.1038/sdata.2016.18.
[56] S.-A. Sansone et al., “FAIRsharing as a community approach to standards, repositories and policies,” *Nature Biotechnology*, vol. 37, pp. 358–367, 2019, doi: 10.1038/s41587-019-0080-8.
[57] Matthias Scheffler *et al.*, “FAIR data enabling new horizons for materials research,” *Nature*, vol. 604, no. 7907, pp. 635–642, 2022, doi: 10.1038/s41586-022-04501-x.
[58] Steven M. Kearnes *et al.*, “The Open Reaction Database,” *Journal of the American Chemical Society*, vol. 143, no. 45, pp. 18820–18826, 2021, doi: 10.1021/jacs.1c09820.
[59] F. Gusev et al., “Machine learning anomaly detection of automated HPLC experiments in the cloud laboratory,” *Digital Discovery*, vol. 4, pp. 3445–3454, 2025, doi: 10.1039/D5DD00253B.
[60] Christian P. Haas *et al.*, “Open-Source Chromatographic Data Analysis for Reaction Optimization and Screening,” *ACS Central Science*, vol. 9, no. 2, pp. 307–317, 2023, doi: 10.1021/acscentsci.2c01042.
[61] Z. Zhou et al., “Localization, inspection, and reasoning module for autonomous workflows in self-driving laboratories,” *Communications Chemistry*, vol. 8, p. 384, 2025, doi: 10.1038/s42004-025-01770-1.
[62] H. Fakhruldeen et al., “Multimodal behaviour trees for robotic laboratory task automation,” in *2025 IEEE International Conference on Robotics and Automation*, 2025, pp. 15872–15878. doi: 10.1109/ICRA55743.2025.11128408.
[63] M. Iovino, E. Scukins, J. Styrud, P. Ögren, and C. Smith, “A survey of behavior trees in robotics and AI,” *Robotics and Autonomous Systems*, vol. 154, p. 104096, 2022, doi: 10.1016/j.robot.2022.104096.
[64] K. Ochiai et al., “Automating care by self-maintainability for full laboratory automation,” *Digital Discovery*, vol. 4, pp. 2285–2297, 2025, doi: 10.1039/D5DD00151J.
[65] M. Leucker and C. Schallhart, “A brief account of runtime verification,” *The Journal of Logic and Algebraic Programming*, vol. 78, no. 5, pp. 293–303, 2009, doi: 10.1016/j.jlap.2008.08.004.
[66] A. Bauer, M. Leucker, and C. Schallhart, “Runtime verification for LTL and TLTL,” *ACM Transactions on Software Engineering and Methodology*, vol. 20, no. 4, pp. 14:1–14:64, 2011, doi: 10.1145/2000799.2000800.
[67] Y. K. Wakabayashi et al., “Bayesian optimization with experimental failure for high-throughput materials growth,” *npj Computational Materials*, vol. 8, p. 180, 2022, doi: 10.1038/s41524-022-00859-8.
[68] R. J. Hickman et al., “Anubis: Bayesian optimization with unknown feasibility constraints for scientific experimentation,” *Digital Discovery*, vol. 4, pp. 2104–2122, 2025, doi: 10.1039/D5DD00018A.
[69] Robert G. Bergman and Rick L. Danheiser, “Reproducibility in Chemical Research,” *Angewandte Chemie International Edition*, vol. 55, no. 41, pp. 12548–12549, 2016, doi: 10.1002/anie.201606591.
[70] L. Lamport, “Time, clocks, and the ordering of events in a distributed system,” *Communications of the ACM*, vol. 21, no. 7, pp. 558–565, 1978, doi: 10.1145/359545.359563.
[71] Michael Walker, Gabriella Pizzuto, Hatem Fakhruldeen, and Andrew I. Cooper, “Go with the flow: deep learning methods for autonomous viscosity estimations,” *Digital Discovery*, vol. 2, no. 5, pp. 1540–1547, 2023, doi: 10.1039/d3dd00109a.
[72] D. B. Rubin, “Inference and missing data,” *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976, doi: 10.1093/biomet/63.3.581.
[73] J. J. Heckman, “Sample selection bias as a specification error,” *Econometrica*, vol. 47, no. 1, pp. 153–161, 1979, doi: 10.2307/1912352.
[74] P. Raccuglia et al., “Machine-learning-assisted materials discovery using failed experiments,” *Nature*, vol. 533, pp. 73–76, 2016, doi: 10.1038/nature17439.
[75] Felix Strieth Kalthoff, Frederik Sandfort, Marius Kühnemund, Felix R. Schäfer, Herbert Kuchen, and Frank Glorius, “Machine Learning for Chemical Reactivity: The Importance of Failed Experiments,” *Angewandte Chemie International Edition*, vol. 61, no. 29, p. e202204647, 2022, doi: 10.1002/anie.202204647.
[76] Anne-Catherine Bédard *et al.*, “Reconfigurable system for automated optimization of diverse chemical reactions,” *Science*, vol. 361, no. 6408, pp. 1220–1225, 2018, doi: 10.1126/science.aat0650.
[77] C. W. Coley et al., “A robotic platform for flow synthesis of organic compounds informed by AI planning,” *Science*, vol. 365, no. 6453, p. eaax1566, 2019, doi: 10.1126/science.aax1566.
[78] Ryan A. Skilton *et al.*, “Remote-controlled experiments with cloud chemistry,” *Nature Chemistry*, vol. 7, no. 1, pp. 1–5, 2015, doi: 10.1038/nchem.2143.
[79] Dario Caramelli *et al.*, “Networking chemical robots for reaction multitasking,” *Nature Communications*, vol. 9, no. 1, p. 3406, 2018, doi: 10.1038/s41467-018-05828-8.
[80] Hatem Fakhruldeen, Gabriella Pizzuto, Jakub Glowacki, and Andrew Ian Cooper, “ARChemist: Autonomous Robotic Chemistry System Architecture,” in *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 2022, pp. 6013–6019. doi: 10.1109/icra46639.2022.9811996.
[81] Joyce Xin-Yan Lim, Dasheng Leow, Quang-Cuong Pham, and Choon-Hong Tan, “Development of a Robotic System for Automatic Organic Chemistry Synthesis,” *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 4, pp. 2185–2190, 2021, doi: 10.1109/tase.2020.3036055.
[82] Benjamin J. Shields *et al.*, “Bayesian reaction optimization as a tool for chemical synthesis,” *Nature*, vol. 590, no. 7844, pp. 89–96, 2021, doi: 10.1038/s41586-021-03213-y.