Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

EMBODIED ARTIFICIAL INTELLIGENCE MUST LEARN LABORATORY CONSTRAINTS BEFORE IT CAN AUTONOMOUSLY CONDUCT PHARMACEUTICAL EXPERIMENTS

Marco De Luca[1]*, Giulia Ferraro[2], Antonio Russo[1], Elena Rossi[3]

1. ***Department of Embodied AI and Laboratory Constraints, Faculty of Pharmacy, University of Naples Federico II, Naples, Italy.***
2. ***Department of Autonomous Experimentation and Physical Reasoning, Faculty of Pharmacy, University of Bologna, Bologna, Italy.***
3. ***Department of Laboratory-Aware AI for Pharma, Faculty of Pharmacy, University of Florence, Florence, Italy.***

ARTICLE INFO

Received:
26 July 2025
Received in revised form:
23 November 2025
Accepted:
30 November 2025
Available online:
28 December 2025

Keywords: Embodied artificial intelligence, Autonomous laboratories, Pharmaceutical experimentation, Laboratory constraints, Procedural grounding, Adaptive execution

ABSTRACT

Artificial intelligence is increasingly connected to automated synthesis, analytical instrumentation, experiment planning, and closed-loop scientific discovery. However, computational competence does not by itself establish the capacity to conduct pharmaceutical experiments. Laboratory work is embodied: every action is conditioned by equipment availability, material identity, container state, spatial reachability, procedural dependencies, timing requirements, contamination controls, hazard boundaries, measurement uncertainty, and human authority. Current autonomous-laboratory architectures demonstrate important elements of planning, robotic execution, sensing, and iterative optimization, yet these capabilities are commonly evaluated separately or summarized through narrow measures of task completion or optimization efficiency. This article develops an original design-principles construct, termed Constraint-Grounded Embodied Laboratory Intelligence, to organize the conditions under which artificial intelligence may progress from generating experimental proposals to participating in bounded pharmaceutical experimentation. The construct represents laboratory intelligence as a continuously updated relation among equipment affordances, material and sample states, executable procedures, temporal commitments, safety envelopes, epistemic provenance, perception, manipulation, adaptive execution, anomaly interpretation, and human takeover. The article argues that no single benchmark, planning score, manipulation-success measure, or model-confidence estimate can establish autonomous laboratory competence. Instead, evaluation must examine whether computational representations remain synchronized with physical laboratory conditions, whether proposed actions satisfy procedural and safety constraints, whether deviations are detected and interpreted appropriately, and whether humans can intervene before consequences become irreversible. The contribution is conceptual rather than empirically validated. Its applicability will depend on laboratory type, pharmaceutical modality, instrumentation, materials, local governance, and prospective evaluation. The proposed framework may nevertheless support more rigorous system design, reporting, comparison, and staged validation of embodied artificial intelligence for pharmaceutical research.

To Cite This Article: De Luca M, Ferraro G, Russo A, Rossi E. Embodied Artificial Intelligence Must Learn Laboratory Constraints before It Can Autonomously Conduct Pharmaceutical Experiments. Pharmacophore. 2025;16(6):66-76. https://doi.org/10.51847/aJmB8aq7Zf

Introduction

Self-driving laboratories are not single algorithms but integrated research systems in which experiment planning, automated execution, sensing, data infrastructure, and scientific oversight must operate as a coordinated whole [1-3]. This distinction is particularly important in pharmaceutical science, where a computationally attractive experiment may still be physically inaccessible, procedurally incomplete, chemically incompatible, temporally invalid, analytically uninterpretable, or unsafe. An autonomous platform therefore cannot be judged only by the quality of its predictions or by how efficiently it searches a design space. It must also be judged by whether it can maintain an accurate relation between digital instructions and the changing physical state of the laboratory.

Corresponding Author: Marco De Luca; Department of Embodied AI and Laboratory Constraints, Faculty of Pharmacy, University of Naples Federico II, Naples, Italy. E-mail: marco.deluca@unina.it

The unresolved problem is that contemporary artificial-intelligence systems generally represent experiments at a level that is more abstract than the conditions governing their execution. Molecular structures, formulation variables, reaction parameters, analytical targets, or optimization objectives may be available to a model, while container geometry, instrument readiness, sample history, calibration status, contamination risk, temporal dependencies, operator authority, and unresolved uncertainty remain external to the decision representation. This separation permits a system to produce a scientifically plausible proposal that cannot be executed as specified or whose execution would not preserve the intended scientific question. Data availability must therefore be distinguished from data suitability, just as model confidence must be distinguished from calibrated uncertainty and computational feasibility from laboratory admissibility.

Integrated robotic platforms show that computational planning can be connected to automated synthesis, purification, and measurement, while closed-loop molecular-discovery systems demonstrate that predicted candidates can be synthesized, evaluated, and returned to the decision process [4, 5]. These achievements establish the feasibility of tightly coupled prediction–execution–measurement cycles within engineered environments. They do not establish that an artificial agent possesses a general understanding of laboratory constraints. In each case, the apparent autonomy remains conditional on human-designed equipment, bounded action vocabularies, available analytical methods, predefined operating ranges, and task-specific representations of what may be attempted.

This article therefore proposes Constraint-Grounded Embodied Laboratory Intelligence (CGELI) as an original design-principles construct for pharmaceutical experimentation. CGELI treats laboratory intelligence as the capacity to select, execute, monitor, interrupt, and document experimental actions while remaining conditioned by an explicit and continuously updated representation of physical, procedural, temporal, safety, epistemic, and human-oversight constraints. The contribution is not presented as a validated architecture, universal ontology, regulatory framework, or deployment-ready system. Its purpose is to clarify which relationships must be represented, how embodied planning differs from abstract optimization, why adaptive execution requires bounded authority, and how future systems should be evaluated across progressively more realistic laboratory settings.

Why Laboratory Intelligence Must Be Physically and Procedurally Grounded

Physical grounding concerns whether an artificial agent represents what can actually be reached, moved, opened, transferred, mixed, heated, measured, cleaned, or isolated in a particular laboratory. Mobile robotic chemists demonstrate that apparently simple experimental sequences depend on navigation, collision avoidance, sample transport, instrument access, task allocation, and continuity of context across physical stations [6, 7]. A plan that refers only to scientific operations such as “prepare,” “react,” or “analyze” conceals the embodied requirements that make those operations possible. Laboratory intelligence must therefore include equipment location, manipulator reach, compatible containers, instrument interfaces, access restrictions, consumable state, calibration status, and the consequences of moving samples between environments.

Procedural grounding concerns whether a scientific instruction has been transformed into operations that are explicit enough to execute and verify. The conversion of synthesis literature into editable chemical code illustrates that narrative procedures must be decomposed into ordered actions, parameters, dependencies, and hardware-compatible commands before automated execution can occur [8]. This requirement is more demanding in pharmaceutical experimentation because procedures frequently contain tacit assumptions about material handling, equilibration, sterility, order of addition, mixing regime, sampling, storage, acceptance criteria, and exception management. A procedure may be syntactically complete while remaining scientifically underdetermined if these assumptions are not represented.

A chemical programming language can connect abstract synthesis instructions to modular robotic operations, showing how executable representations mediate between scientific intent and physical hardware [9]. Yet procedural executability should not be confused with procedural validity. A robot may complete every encoded action while using an unsuitable vessel, an unstable material, an inappropriate analytical method, or a timing sequence that changes the interpretation of the outcome. Grounding must therefore join procedural structure with the current material state, equipment affordances, measurement context, and explicit checkpoints at which the system verifies that prerequisites remain satisfied.

Tool-using artificial-intelligence agents may coordinate information retrieval, planning services, code execution, and laboratory interfaces, but access to tools does not establish robust physical understanding, unrestricted safety, or general competence [10]. CGELI therefore proposes that every candidate action should be evaluated through a constraint field before it becomes executable. The field includes physical reachability, procedural prerequisites, material continuity, timing validity, hazard conditions, epistemic uncertainty, and the authority required for continuation. These constraints are not secondary filters applied after optimization; they define the action space within which optimization is permitted. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why laboratory intelligence must be physically and procedurally grounded within the article’s central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for Why laboratory intelligence must be physically and procedurally grounded

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement

Physical reachability and equipment affordances	Can the proposed action be performed by the available robot, instrument, vessel, and laboratory layout?	Mobile robotic systems show that navigation, transport, station access, and instrument operation are integral to experimental execution [6].	Establishes that laboratory actions must be represented as physically situated operations rather than abstract scientific verbs.	A planner may generate an action that is chemically reasonable but unreachable, mechanically incompatible, or unsupported by the configured equipment.	Declared equipment capability does not establish availability, calibration, suitability, or safe use for a specific pharmaceutical material.
Cross-station task continuity	Can sample identity, task context, and procedural state be preserved across transfers among synthesis and analytical stations?	Multi-robot laboratory systems demonstrate the need to coordinate physical transfers and retain workflow context across instruments [7].	Connects embodiment to chain of custody, routing, synchronization, and state persistence.	A sample may be physically transferred while its identity, history, or required next action becomes ambiguous.	Successful transport does not establish sample integrity, analytical validity, or scientific comparability.
Executable procedural representation	Has narrative scientific intent been converted into explicit actions, parameters, dependencies, and verification points?	Digitization systems show that published procedures require formal representation before robotic execution [8].	Defines procedural grounding as a prerequisite for autonomous action.	Natural-language ambiguity or omitted tacit knowledge may be converted into falsely precise instructions.	Machine-readable syntax does not establish chemical correctness, completeness, reproducibility, or safety.
Hardware-compatible action semantics	Does every encoded action map to an available and validated unit operation?	Chemical programming demonstrates the mapping of abstract synthesis actions to modular hardware functions [9].	Links procedural logic to concrete laboratory affordances.	A formally valid command may invoke an operation outside the platform's validated material, volume, temperature, or containment envelope.	Executability is platform-specific and cannot be assumed to transfer across laboratories or equipment configurations.
Material-state continuity	Does the system know which material is present, where it is located, how it was prepared, and whether it remains suitable for the next action?	Robotic execution requires persistent links among reagents, containers, actions, and observations [6].	Extends grounding from equipment state to samples, intermediates, and chain of custody.	Incorrect identity, contamination, degradation, or unrecorded preparation history may invalidate later actions.	Digital identity does not establish purity, stability, sterility, biological suitability, or pharmaceutical quality.
Tool-mediated planning	Does the agent understand the operational implications of invoking a tool, or does it merely generate a compatible command?	Tool-using chemical agents demonstrate bounded coordination of software and laboratory functions [10].	Distinguishes interface access from embodied laboratory understanding.	Fluent planning may conceal missing preconditions, inappropriate escalation, or unsupported assumptions about physical outcomes.	Tool access and task completion do not establish general laboratory intelligence, unrestricted autonomy, or deployment readiness.
Constraint-field integration	Are physical, procedural, material, temporal, hazard, epistemic, and authority constraints evaluated together before execution?	This article synthesizes the relationships supported individually across robotic, procedural, and agent-based systems.	Provides the conceptual bridge from demonstrated components to the proposed CGELI construct.	The framework could be misread as an empirically complete ontology or validated control architecture.	The constraint field is a proposed organizing construct that requires laboratory-specific definition and prospective validation.

Equipment, Materials, Protocols, Hazards, and Temporal Constraints

Equipment constraints determine not only whether an operation exists in principle but whether it can be performed by a specific configuration at a specific time. Reconfigurable automated platforms show that reaction optimization depends on the deliberate arrangement of reagent delivery, reactors, separators, analytical instruments, and admissible operating ranges [11]. In an

embodied pharmaceutical system, equipment state would therefore include installed modules, current configuration, calibration, cleaning status, occupied capacity, maintenance condition, compatible vessels, accessible ranges, and interlocks. Materials must be represented in relation to these affordances. A powder, viscous formulation, volatile solvent, photosensitive intermediate, biological sample, or sterile preparation may require different containment, mixing, transfer, environmental, and measurement conditions even when the computational action is described by the same verb.

Material and protocol constraints are inseparable because each procedural action changes the identity, location, condition, or evidentiary status of a physical sample. Robotic searches for chemical reactivity demonstrate that material combinations, preparation actions, analytical signals, and subsequent experiment choices can be joined in a machine-readable loop [12]. For pharmaceutical experimentation, the required state would be broader: sample identity, batch origin, quantity, concentration, physical form, container, temperature, age, exposure history, contamination status, prior manipulations, analytical results, and unresolved discrepancies may all determine whether the next action is admissible. The protocol should consequently be represented as a dependency graph in which actions have explicit prerequisites, verification points, permitted deviations, and consequences for sample validity. An action should not be considered complete merely because a device moved; completion should require evidence that the intended material transformation or measurement condition was achieved.

Temporal constraints convert the laboratory from a static collection of capabilities into a state-transition system. Multi-step autonomous fluidic experimentation demonstrates that sequential operations require persistent intermediate states, reusable control actions, and decision logic that respects the order of execution [13]. Pharmaceutical workflows add incubation periods, dissolution and equilibration windows, sample-stability limits, instrument queues, cleaning cycles, time-dependent degradation, delayed measurements, synchronization among parallel operations, and deadlines after which a sample or result may no longer support the intended interpretation. CGELI therefore proposes a temporal commitment layer that records not only expected duration but also earliest and latest permissible actions, dependencies among processes, uncertainty in completion time, and the consequences of delay. Hazards should be treated as cross-cutting constraints over this temporal and material state: incompatible combinations, loss of containment, unsafe temperatures, pressure excursions, contamination, or expired intervention windows must be able to suspend optimization and transfer authority to a safer control state. This proposed integration does not establish that all hazards can be represented in advance; unknown interactions, hidden physical conditions, and sensor failure remain unavoidable boundary conditions.

Embodied Representation of Pharmaceutical Laboratory State

An embodied laboratory agent requires more than a database of instruments and samples; it requires an operational state that can be queried before, during, and after every action. Laboratory orchestration architectures show the importance of coordinating experiment requests, instruments, data flows, and decision services through a common control layer [14]. In CGELI, this orchestration layer is extended into a continuously updated belief state comprising equipment affordances, material identity and condition, procedural position, temporal commitments, hazard status, epistemic uncertainty, provenance, and human authority. The term *belief state* is deliberate: the representation is an evidence-conditioned estimate of the laboratory, not a claim that every physically relevant variable is known.

Structured reaction-data infrastructures demonstrate how inputs, conditions, operations, outcomes, and provenance can be represented in interoperable schemas [15]. Pharmaceutical laboratory state, however, must also preserve relations that may not be captured by reaction records alone: a sample's container and location, environmental exposure, preparation history, stability window, chain of custody, pending quality checks, instrument suitability, and whether a result is observed, inferred, predicted, or disputed. CGELI therefore proposes that state variables carry source, timestamp, confidence, validation status, and dependency information. Contradictions should remain visible rather than being silently reconciled, because a disagreement between inventory, vision, telemetry, and analytical evidence may itself be the condition that requires suspension.

Procedural language can be converted into typed and ordered synthesis actions, supporting the representation of protocols as action graphs rather than unstructured narrative alone [16]. Within the proposed construct, each node in such a graph has preconditions, required resources, expected state transitions, verification criteria, permitted deviations, and recovery implications. This procedural layer is connected to a temporal ledger that records duration, deadlines, synchronization, stability windows, and delayed effects; to an equipment-affordance layer that identifies executable operations; and to a material layer that preserves identity and condition. The resulting representation does not merely record what the protocol says should happen. It asks whether the evidence available at that moment supports the conclusion that the required state has been reached.

Modular, object-oriented workflow abstractions can facilitate the reuse and evolution of automated laboratory processes across platforms [17]. Nevertheless, autonomous systems that integrate computational priors, robotic execution, and characterization remain dependent on task-specific equipment, domain rules, and decision representations [18]. CGELI consequently separates reusable component definitions from laboratory-specific instantiation: the architecture may be portable in principle, whereas its equipment models, material semantics, hazard rules, temporal parameters, and authority boundaries must be locally validated. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop embodied representation of pharmaceutical laboratory state within the article's central argument.

Table 2. Components, relationships, uncertainties, and validation needs within Embodied representation of pharmaceutical laboratory state

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Equipment-affordance layer	Instrument registry, calibration and maintenance state, spatial layout, consumables, interfaces, operating ranges, and interlocks	Determine which actions the configured laboratory can physically perform	Set of currently executable, conditionally executable, and prohibited equipment actions	Vendor-specific behavior, undocumented settings, drift, temporary unavailability, or incomplete capability descriptions	Laboratory-specific capability tests and state-synchronization assessment under normal and perturbed conditions
Material identity-and-condition layer	Sample identifiers, amounts, concentrations, containers, locations, temperatures, ages, exposure histories, and contamination indicators	Maintain continuity between digital entities and physical samples	Versioned material state and chain-of-custody record	Sensors may not uniquely establish composition, purity, degradation, or contamination	Orthogonal identity checks, provenance audits, and material-specific stability studies
Procedural dependency layer	Structured protocol, action types, prerequisites, parameters, branches, verification points, and recovery rules	Convert scientific intent into auditable executable dependencies	Current protocol position and admissible next actions	Tacit knowledge and laboratory-specific judgment may be absent from the encoded procedure	Comparison of written, encoded, and expert-performed protocols across exception scenarios
Temporal commitment layer	Action durations, incubation periods, stability windows, queues, synchronization requirements, and delayed effects	Prevent individually valid actions from forming an invalid schedule	Time-bounded action permissions, deadlines, and escalation conditions	Durations and material stability may be stochastic or context-dependent	Prospective validation against time-stamped laboratory operations and sample-quality evidence
Hazard and safety-envelope layer	Material hazards, incompatible states, containment requirements, exposure controls, equipment limits, and institutional rules	Place non-negotiable constraints above optimization objectives	Forbidden states, constrained operating regions, mandatory interlocks, and escalation triggers	Unknown interactions, hidden hazards, and sensor blind spots cannot be fully enumerated	Independent risk assessment, fault injection, interlock testing, and local safety approval
Epistemic and provenance layer	Measurement quality, source identity, model version, uncertainty, lineage, contradictions, and missing data	Separate observation, inference, prediction, and unresolved unknowns	Auditable evidence state with explicit uncertainty and disagreement	Model confidence may be miscalibrated, and complete logs may still document an invalid experiment	Calibration studies, lineage audits, contradiction tests, and expert adjudication
Perception–action state estimator	Vision, robot telemetry, instrument status, analytical data, inventory, environmental sensing, and protocol progress	Fuse multimodal observations into an actionable laboratory belief state	Updated state estimate with confidence, freshness, and unresolved discrepancies	Occlusion, latency, correlated sensor failure, and unmeasured variables can corrupt the estimate	Matched physical ground truth, perturbation tests, stale-state detection, and abstention assessment
Constraint-aware planner	Scientific objective, current state, feasible-action models, uncertainty, and temporal and safety rules	Propose informative actions only after checking declared prerequisites and constraints	Ranked admissible actions with rationale and unresolved assumptions	Objectives may be incomplete or encourage proxy optimization	Adversarial constraint cases, comparison with expert-reviewed plans, and explanation audits
Adaptive execution, anomaly, and takeover layer	Approved plan, live state updates, deviation signals, recovery permissions, and authority model	Verify preconditions, monitor execution, adapt within bounds, and pause or transfer control when limits are exceeded	Verified completion, bounded recovery, stop, quarantine, or human handover	Recovery may alter scientific meaning; anomalies may be undiagnosed; humans may lack time or context	Fault-injection studies, protocol-equivalence review, timed takeover experiments, and audit-trail verification

Figure 1 presents the embodied pharmaceutical laboratory constraint field, showing how the article's key components and boundaries are connected within embodied representation of pharmaceutical laboratory state.

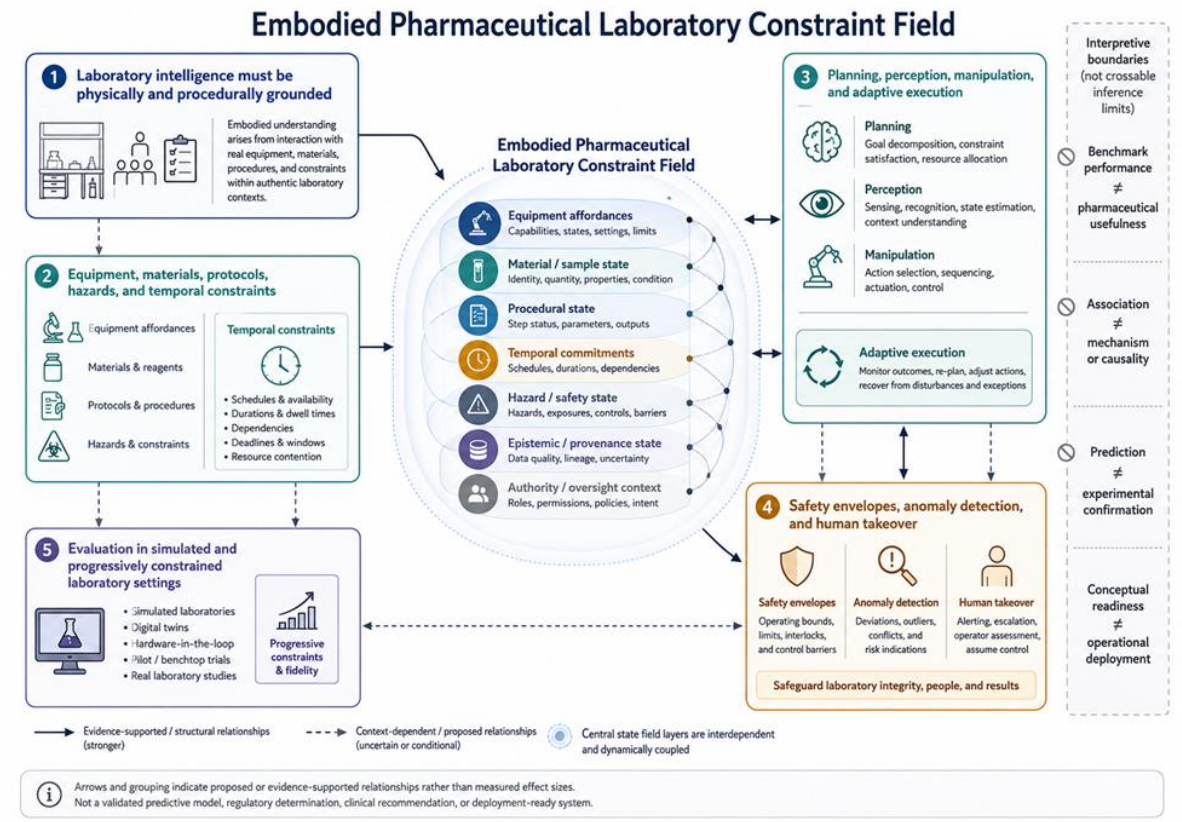

Figure 1. Embodied Pharmaceutical Laboratory Constraint Field

The figure is an original conceptual synthesis that organizes the article's central contribution across laboratory intelligence must be physically and procedurally grounded, equipment, materials, protocols, hazards, and temporal constraints, temporal constraints, embodied representation of pharmaceutical laboratory state, planning, perception, manipulation, and adaptive execution, adaptive execution. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Planning, Perception, Manipulation, and Adaptive Execution

Planning becomes embodied when a proposed scientific action is represented together with the observations, physical resources, and manipulations required to complete it. An AI-driven robotic chemist can coordinate synthesis planning with reagent handling and batch operations, illustrating that execution depends on translation from a chemical objective to a concrete sequence of physical actions [19]. CGELI extends this relation by requiring every planned action to include its preconditions, expected state transition, observation plan, completion evidence, uncertainty, and permissible recovery. The planner should therefore return not only a preferred experiment but also an executable argument for why that experiment is currently admissible.

Perception provides the evidence by which the system determines whether those preconditions and expected transitions are satisfied. Sensor-guided robotic chemistry demonstrates the value of real-time analytical feedback, multistep execution, and explicit cleaning or reset operations during autonomous exploration [20]. For pharmaceutical systems, perception may combine robot telemetry, machine vision, mass or volume checks, environmental monitoring, analytical measurements, inventory state, and protocol progress. No single stream should be treated as definitive when consequences are substantial. Agreement, disagreement, freshness, and provenance must be represented so that the agent can proceed, remeasure, abstain, quarantine a sample, or request human review.

Manipulation quality must be defined scientifically rather than mechanically. A gripper may successfully move a vial while compromising its identity; a liquid handler may complete a transfer while introducing carryover; a mixer may rotate at the commanded setting without producing the required material state. Self-driving platforms that couple model proposals, robotic processing, measurement, and iterative updating show why planned variables must remain consistently linked to physical process parameters and observed results [21]. CGELI therefore treats manipulation as a verified state transition. Completion requires evidence that the intended sample, container, quantity, condition, and destination are consistent with the protocol and that no unresolved event has invalidated the experiment.

Adaptive execution closes the loop but must not become unconstrained improvisation. Real-time active learning can select subsequent experiments from live measurements and uncertainty, demonstrating the value of updating decisions as evidence accumulates [22]. In the proposed framework, adaptation is permitted only within pre-authorized scientific and safety bounds. The controller may adjust scheduling, repeat a measurement, select an approved alternative instrument, or execute a validated recovery, but it may not silently redefine the experimental objective, ignore failed controls, relax a hazard constraint, or treat completion as more important than validity. When the required adaptation would change the scientific meaning or risk class of the experiment, authority must pass to a human decision point.

Safety Envelopes, Anomaly Detection, and Human Takeover

Safety cannot be appended to an autonomous system as a final checklist. Safe-learning research shows that learning-enabled control requires explicit constraints, uncertainty handling, monitoring, and mechanisms that prevent performance optimization from overriding safety requirements [23]. In CGELI, a safety envelope is a dynamic set of admissible states and transitions defined by material hazards, equipment limits, containment, environmental conditions, procedure stage, and available intervention. Hard constraints prohibit actions regardless of expected scientific value. Conditional constraints permit actions only when evidence and safeguards are present. Unknown conditions trigger conservative behavior rather than being treated as evidence of safety.

The safety envelope must also define the relation between robots and people. Human–robot collaboration research emphasizes both physical safeguards and interfaces through which operators can understand, interrupt, and redirect robotic activity [24]. Pharmaceutical laboratories add chemical exposure, contamination control, sterility, potent compounds, sensitive biological materials, and quality-system considerations that are not captured by generic collision avoidance. Human access zones, robot speed, containment status, ventilation, sample risk, and procedural consequence may therefore alter one another dynamically. The system should expose these dependencies in a form that supports intervention rather than merely reporting an opaque risk score.

Anomaly detection supplies warnings when observed behavior departs from an expected state or trajectory, but its meaning is conditional on data, model assumptions, and the definition of normality. Reviews of deep anomaly detection show substantial variation in assumptions and performance under sparse labels, high dimensionality, and distribution shift [25]. An abnormal signal may reflect a material transformation, sensor drift, instrument failure, contamination, procedural error, or a benign condition omitted from training. CGELI consequently distinguishes anomaly detection from anomaly interpretation. A score may justify pause or additional measurement, but it does not establish cause, prescribe a safe recovery, or demonstrate that undetected operation is safe.

Human takeover is therefore an engineered control transition, not a rhetorical assurance that an operator remains "in the loop." Human-automation evidence indicates that operators can lose situation awareness, miscalibrate trust, or be unable to intervene effectively when control is returned unexpectedly [26]. The proposed authority layer specifies who may stop, pause, approve, modify, quarantine, resume, or terminate an experiment; what information must accompany a handover; and how much time remains before consequences become irreversible. Interfaces should communicate current material and equipment state, the triggering evidence, uncertainty, actions already taken, prohibited responses, and safe options. Human presence alone is insufficient: takeover performance must be trained, timed, observed, and evaluated.

Evaluation in Simulated and Progressively Constrained Laboratory Settings

Evaluation should begin from the recognition that an autonomous laboratory is a coupled system, not an optimization algorithm operating in isolation. Cross-domain comparisons of Bayesian optimization show that planning performance varies with the experimental landscape and that conclusions derived from one task may not generalize to another [27]. For embodied pharmaceutical AI, evaluation must therefore cover state completeness, synchronization, constraint satisfaction, perception robustness, manipulation integrity, adaptive execution, anomaly response, takeover effectiveness, scientific validity, and transfer across settings. A system can perform strongly on one dimension while remaining unsuitable on another.

Noisy experimental emulators can support controlled comparison of experiment-planning methods and can expose sensitivity to noise, objective structure, and limited observations [28]. They are appropriate for early testing of planner behavior, protocol logic, temporal scheduling, and declared constraint handling. Their limitations must remain explicit. An emulator cannot reveal every unmodeled mechanical interaction, contamination pathway, sensor artifact, material instability, interface failure, or human coordination problem. Success in an emulator should authorize only the next evaluation stage, not a conclusion that physical laboratory competence has been demonstrated.

Simulation should be assessed by whether it predicts comparative performance in the physical system. Sim2Real research has shown the importance of testing whether method rankings and conclusions obtained in simulation persist in real-world trials [29]. CGELI therefore proposes matched tasks across offline protocol checking, noisy emulation, physical mock-ups with inert materials, low-hazard constrained laboratory operations, and domain-relevant pharmaceutical experiments under human authority. Each transition should seek new failure modes as well as retained performance. A decrease in performance is informative when it identifies missing state variables, invalid dynamics, inadequate sensing, or overfitting to simulator assumptions.

Real-world learning differs from idealized benchmarks because observations are limited, actions may be delayed or irreversible, conditions may change, and safety constraints can dominate exploration [30]. Progressive evaluation should

introduce these properties deliberately rather than waiting for them to appear in consequential experiments. Advancement should require evidence that the digital state remains aligned with the physical system, forbidden actions are rejected, permitted actions are executed without compromising sample validity, anomalies trigger appropriate escalation, and humans can take over effectively. **Table 3** organizes the evidence, constructs, and interpretation boundaries needed to develop evaluation in simulated and progressively constrained laboratory settings within the article's central argument.

Table 3. Evaluation, implementation, and research priorities arising from Embodied Artificial Intelligence Must Learn Laboratory Constraints before It Can Autonomously Conduct

Evaluation dimension	What must be tested	Suitable evidence or assessment	Failure signal	Interpretive limitation	Decision relevance
State completeness and accuracy	Whether all variables required to determine action admissibility are represented and current	Ground-truth instrument, material, protocol, temporal, and hazard annotations; contradiction and stale-state tests	The system acts despite an omitted prerequisite, unresolved contradiction, or outdated state	High coverage in a test set does not prove completeness in an open laboratory	Determines whether planning may proceed or must abstain
Constraint satisfaction	Whether proposed actions comply with equipment, procedure, timing, hazard, and authority rules	Adversarial valid and invalid plans; independent expert adjudication; explanation audit	A high-value action is selected outside a declared constraint, or a valid action is repeatedly rejected	Passing known constraints cannot reveal unknown constraints	Supports planner revision and defines whether progression is permissible
Perception robustness	Whether the belief state remains reliable under noise, latency, drift, occlusion, and sensor disagreement	Matched nominal and perturbed scenarios with physical ground truth and abstention assessment	Plausible but incorrect sensor fusion is accepted without remeasurement or escalation	Test perturbations may not reproduce all material-specific artifacts	Determines sensing redundancy and escalation requirements
Manipulation and sample integrity	Whether physical actions preserve identity, quantity, containment, cleanliness, and scientifically relevant material condition	Repeated trials across liquids, powders, viscous materials, containers, and analytical handoffs	Mechanical completion occurs with contamination, loss, misidentification, or invalid material state	Generic robotic success may ignore pharmaceutical quality attributes	Defines the validated task and material envelope
Adaptive execution and recovery	Whether adaptation remains inside approved bounds and preserves the protocol's scientific meaning	Fault injection with predefined permitted and forbidden adaptations; audit-trail review	Recovery conceals failure, changes the experiment, or creates a new hazard	Completion after recovery does not prove equivalence with the intended procedure	Determines when autonomous recovery is allowed and when human approval is mandatory
Anomaly and hazard response	Whether unsafe or scientifically invalid trajectories are detected early enough for effective action	Near-miss libraries, distribution-shift scenarios, causal fault labels where available, and time-to-alert analysis	Rare or correlated faults appear normal, or false alarms make oversight ineffective	Detection does not establish causal diagnosis or safe corrective action	Sets pause, quarantine, shutdown, and investigation rules
Human takeover	Whether an operator can understand and control the situation before consequences become irreversible	Timed handover studies under realistic workload, uncertainty, and incomplete information	The operator lacks context, authority, time, or a safe action pathway	Performance depends on role, expertise, fatigue, interface, and training	Defines staffing, interface, alarm, and authority requirements
Scientific validity and provenance	Whether successful execution preserves the scientific question, controls, lineage, and interpretability	Replication, control-condition review, blinded expert assessment, and complete action and data lineage	A technically complete experiment cannot answer the intended question or cannot be reconstructed	Reproducible execution is not equivalent to mechanism, therapeutic value, or regulatory acceptability	Determines whether results can support further scientific decisions
Progressive transfer	Whether findings persist from emulator to mock-up and	Matched tasks, rank preservation, sim-to-real	A policy exploits simulator assumptions or fails when	Passing one stage does not authorize movement to a	Supports independent advancement,

	increasingly realistic physical settings	predictivity, performance degradation, and new-failure discovery	unmodeled physical details appear	higher-consequence stage	rollback, and redesign decisions

Figure 2 presents the evidence, validation, and translation boundary observatory for embodied artificial intelligence must learn laboratory constraints, showing how the article's key components and boundaries are connected within evaluation in simulated and progressively constrained laboratory settings.

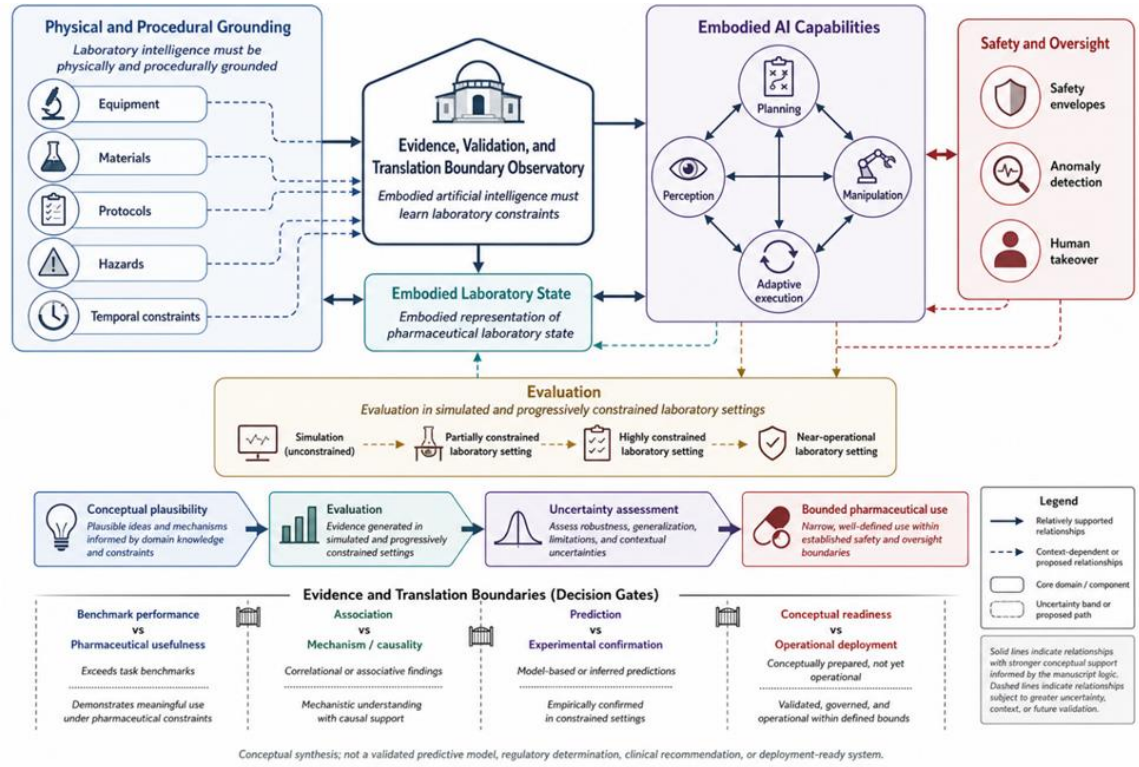

Figure 2. Evidence, Validation, and Translation Boundaries

The figure is an original conceptual synthesis that shows how claims move from conceptual plausibility through evaluation, uncertainty assessment, and bounded pharmaceutical use. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Limitations and Boundary Conditions

CGELI is a proposed conceptual architecture and vocabulary, not a completed software specification or experimentally established theory of laboratory intelligence. Its layers may help expose dependencies that are hidden when planning, robotics, data management, and safety are assessed separately, but their completeness cannot be assumed. Laboratories differ in equipment, layout, materials, procedures, analytical methods, hazard classifications, staffing, quality systems, and acceptable uncertainty. A representation sufficient for one workflow may omit decisive variables in another, and an ontology detailed enough to describe a laboratory can still fail to remain synchronized with its physical state.

The supporting evidence is heterogeneous and includes robotic chemistry, materials acceleration, data infrastructure, safe robotics, anomaly detection, human-automation research, and simulation-to-reality evaluation. Autonomous materials platforms demonstrate sophisticated integration while remaining domain- and platform-specific [18]. Principles transferred from industrial human–robot collaboration require separate validation for chemical exposure, contamination, sterile operations, and pharmaceutical quality [24]. The article therefore does not claim that evidence from one scientific domain proves pharmaceutical applicability. It uses that evidence to identify design requirements and testable boundaries.

Misuse could arise if the construct were treated as a certification checklist or if individual measures were substituted for system validity. An anomaly score is not a causal diagnosis or guarantee of safe operation [25]. Likewise, strong comparative performance in simulation does not alone establish physical-system usefulness [29]. Completion of a robotic action does not prove sample integrity; execution of a protocol does not prove scientific validity; predictive confidence does not prove calibrated uncertainty; and human presence does not prove effective oversight. CGELI cannot establish clinical utility,

manufacturing suitability, regulatory acceptability, universal applicability, or readiness for unsupervised pharmaceutical experimentation.

Research Agenda and Implementation Priorities

The first priority is a laboratory-constraint ontology that joins equipment capabilities, sample condition, executable actions, temporal dependencies, hazards, provenance, and authority without erasing local specificity. Structured reaction schemas provide a useful foundation for representing experimental inputs, operations, and outcomes [15]. Action-extraction methods similarly show how procedural prose can be converted into ordered operations while revealing the risk that omitted tacit knowledge becomes hidden behind formal syntax [16]. Future studies should compare written, encoded, and expert-performed protocols; annotate exceptions and recovery decisions; and test whether independent laboratories assign compatible meanings to shared state variables.

The second priority is empirical evaluation of embodied validity rather than generic robotic success. Shared tasks should include liquids, powders, viscous formulations, heterogeneous samples, diverse containers, cleaning and analytical handoffs, and deliberately introduced state conflicts. Manipulation studies should measure identity preservation, loss, contamination, material-condition change, and evidence quality. Anomaly research should use causal fault libraries and correlated failures, while takeover studies should examine whether operators receive sufficient context and time. Human–robot interface evidence supports the need for understandable intervention mechanisms [24]. Human-automation research further indicates that situation awareness and trust must be designed rather than presumed [26].

Implementation should proceed through staged, reversible commitments. Offline protocol checking and constraint simulation should precede noisy emulators, inert-material mock-ups, low-hazard physical tasks, and domain-relevant pharmaceutical work under explicit human authority. Cross-domain benchmark studies can inform comparative evaluation design but should not supply universal pass thresholds [27]. Matched simulation and physical tasks should test whether rankings, failure modes, and control conclusions transfer [29]. Multi-laboratory replication, independent gate review, auditability, model-change control, and predeclared rollback criteria are essential. The near-term goal is not unrestricted autonomy; it is to establish where bounded assistance is reliable, where uncertainty requires abstention, and what evidence would justify a carefully controlled increase in authority.

Conclusion

Autonomous pharmaceutical experimentation requires a form of intelligence that is inseparable from the laboratory in which it acts. The proposed CGELI construct reframes autonomy as a continuously constrained relation among equipment, materials, procedures, time, hazards, evidence, perception, manipulation, adaptive control, anomaly interpretation, and human authority. Its central implication is that no prediction score, optimization result, robotic completion rate, or confidence estimate can establish laboratory competence alone. A system must demonstrate that its digital state remains aligned with physical conditions, that its actions preserve scientific and material validity, that adaptation remains within authorized bounds, and that uncertainty can produce pause, abstention, rollback, or effective human takeover. CGELI is not validated, universal, or deployment-ready, but it offers a disciplined basis for designing and testing embodied artificial intelligence without confusing computational capability with pharmaceutical usefulness. Progress should therefore be measured by the quality of constraint representation, the discovery and management of failure, and the preservation of accountable scientific judgment as experimental authority increases.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Tom G, Schmid SP, Baird SG, Cao Y, Darvish K, Hao H, et al. Self-driving laboratories for chemistry and materials science. Chem Rev. 2024;124(16):9633-732. doi:10.1021/acs.chemrev.4c00055
2. Häse F, Roch LM, Aspuru-Guzik A. Next-generation experimentation with self-driving laboratories. Trends Chem. 2019;1(3):282-91. doi:10.1016/j.trechm.2019.02.007
3. Stach EA, DeCost BL, Kusne AG, Hattrick-Simpers J, Brown KA, Reyes KG, et al. Autonomous experimentation systems for materials development: A community perspective. Matter. 2021;4(9):2702-26. doi:10.1016/j.matt.2021.06.036
4. Coley CW, Thomas DA 3rd, Lummiss JAM, Jaworski JN, Breen CP, Schultz V, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science. 2019;365(6453). doi:10.1126/science.aax1566

5. Koscher BA, Canty RB, McDonald MA, Greenman KP, McGill CJ, Bilodeau CL, et al. Autonomous, multiproperty-driven molecular discovery: From predictions to measurements and back. Science. 2023;382(6677). doi:10.1126/science.adi1407
6. Burger B, Maffettone PM, Gusev VV, Aitchison CM, Bai Y, Wang X, et al. A mobile robotic chemist. Nature. 2020;583(7815):237-41. doi:10.1038/s41586-020-2442-2
7. Dai T, Vijayakrishnan S, Szczypiński FT, Ayme JF, Simaei E, Fellowes T, et al. Autonomous mobile robots for exploratory synthetic chemistry. Nature. 2024;635(8040):890-7. doi:10.1038/s41586-024-08173-7
8. Mehr SHM, Craven M, Leonov AI, Keenan G, Cronin L. A universal system for digitization and automatic execution of the chemical synthesis literature. Science. 2020;370(6512):101-8. doi:10.1126/science.abc2986
9. Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, et al. Organic synthesis in a modular robotic system driven by a chemical programming language. Science. 2019;363(6423). doi:10.1126/science.aav2211
10. Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. Nature. 2023;624(7992):570-8. doi:10.1038/s41586-023-06792-0
11. Bédard AC, Adamo A, Aroh KC, Russell MG, Bedermann AA, Torosian J, et al. Reconfigurable system for automated optimization of diverse chemical reactions. Science. 2018;361(6408):1220-5. doi:10.1126/science.aat0650
12. Granda JM, Donina L, Dragone V, Long DL, Cronin L. Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature. 2018;559(7714):377-81. doi:10.1038/s41586-018-0307-8
13. Volk AA, Epps RW, Yonemoto DT, Masters BS, Castellano FN, Reyes KG, et al. AlphaFlow: Autonomous discovery and optimization of multi-step chemistry using a self-driven fluidic lab guided by reinforcement learning. Nat Commun. 2023;14(1):1403. doi:10.1038/s41467-023-37139-y
14. Roch LM, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker LPE, Hein JE, et al. ChemOS: Orchestrating autonomous experimentation. Sci Robot. 2018;3(19). doi:10.1126/scirobotics.aat5559
15. Kearnes SM, Maser MR, Wleklinski M, Kast A, Doyle AG, Dreher SD, et al. The Open Reaction Database. J Am Chem Soc. 2021;143(45):18820-6. doi:10.1021/jacs.1c09820
16. Vaucher AC, Zipoli F, Geluykens J, Nair VH, Schwaller P, Laino T. Automated extraction of chemical synthesis actions from experimental procedures. Nat Commun. 2020;11(1):3601. doi:10.1038/s41467-020-17266-6
17. Leong CJ, Low KYA, Recatala-Gomez J, Quijano Velasco P, Vissol-Gaudin E, Tan JD, et al. An object-oriented framework to enable workflow evolution across materials acceleration platforms. Matter. 2022;5(10):3124-34. doi:10.1016/j.matt.2022.08.017
18. Szymanski NJ, Rendy B, Fei Y, Kumar RE, He T, Milsted D, et al. An autonomous laboratory for the accelerated synthesis of novel materials. Nature. 2023;624(7990):86-91. doi:10.1038/s41586-023-06734-w
19. Ha T, Lee D, Kwon Y, Park MS, Lee S, Jang J, et al. AI-driven robotic chemist for autonomous synthesis of organic molecules. Sci Adv. 2023;9(44). doi:10.1126/sciadv.adj0461
20. Jiang Y, Salley D, Sharma A, Keenan G, Mullen M, Cronin L. An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials. Sci Adv. 2022;8(40). doi:10.1126/sciadv.abo2626
21. MacLeod BP, Parlane FGL, Morrissey TD, Häse F, Roch LM, Dettelbach KE, et al. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv. 2020;6(20). doi:10.1126/sciadv.aaz8867
22. Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun. 2020;11(1):5966. doi:10.1038/s41467-020-19597-w
23. Brunke L, Greeff M, Hall AW, Yuan Z, Zhou S, Panerati J, et al. Safe learning in robotics: From learning-based control to safe reinforcement learning. Annu Rev Control Robot Auton Syst. 2022;5:411-44. doi:10.1146/annurev-control-042920-020211
24. Villani V, Pini F, Leali F, Secchi C. Survey on human-robot collaboration in industrial settings: Safety, intuitive interfaces and applications. Mechatronics. 2018;55:248-66. doi:10.1016/j.mechatronics.2018.02.009
25. Pang G, Shen C, Cao L, van den Hengel A. Deep learning for anomaly detection: A review. ACM Comput Surv. 2021;54(2):38. doi:10.1145/3439950
26. Endsley MR. From here to autonomy: Lessons learned from human-automation research. Hum Factors. 2017;59(1):5-27. doi:10.1177/0018720816681350
27. Liang Q, Gongora AE, Ren Z, Tiihonen A, Liu Z, Sun S, et al. Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains. NPJ Comput Mater. 2021;7(1):188. doi:10.1038/s41524-021-00656-9
28. Häse F, Aldeghi M, Hickman RJ, Roch LM, Christensen M, Liles E, et al. Olympus: A benchmarking framework for noisy optimization and experiment planning. Mach Learn Sci Technol. 2021;2(3):035021. doi:10.1088/2632-2153/abedc8
29. Kadian A, Truong J, Gokaslan A, Clegg A, Wijmans E, Lee S, et al. Sim2Real predictivity: Does evaluation in simulation predict real-world performance? IEEE Robot Autom Lett. 2020;5(4):6670-7. doi:10.1109/LRA.2020.3013848
30. Dulac-Arnold G, Levine N, Mankowitz DJ, Li J, Paduraru C, Gowal S, et al. Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis. Mach Learn. 2021;110(9):2419-68. doi:10.1007/s10994-021-05961-4