Agentic AI for Computational Pharmacology: Orchestrating Molecular Search, Evidence Appraisal, Mechanistic Reasoning, and Candidate Prioritization

Lina Hassan[1]*, Omar Khalaf[2], Reem Jaber[1]

[1]Department of AI for Dead Sea and Desert Plant Adaptogens, Faculty of Pharmacy, University of Jordan, Amman, Jordan.

[2]Department of Machine Learning for Kinase Inhibition, Faculty of Pharmacy, Jordan University of Science and Technology, Irbid, Jordan.

ABSTRACT

Agentic artificial intelligence is emerging as a potential approach for coordinating complex computational research workflows that require multiple searches, tools, reasoning stages, and review checkpoints. In computational pharmacology, such coordination may support the integration of molecular search, literature retrieval, evidence appraisal, target and pathway interpretation, safety assessment, and candidate prioritization. This article proposes a conceptual agentic AI framework designed to organize these activities while preserving source grounding, pharmacological caution, auditability, and human decision authority. The framework separates molecular and evidence-retrieval functions from source-quality appraisal, mechanistic reasoning, safety review, and multi-criteria prioritization. It further introduces explicit controls for tool permissions, evidence provenance, citation traceability, conflict detection, uncertainty communication, expert escalation, and validation planning. Molecular search outputs are treated as research hypotheses rather than validated candidates, while mechanistic interpretations are framed as evidence-linked plausibility assessments rather than confirmation of biological mechanism. Candidate rankings are similarly restricted to provisional decision-support outputs that require pharmacological, medicinal-chemistry, toxicological, and experimental review. The principal contribution is a structured architecture in which specialized agents can assist task decomposition and evidence organization without independently establishing biological truth, safety, efficacy, translational readiness, or clinical utility. By embedding human oversight and validation boundaries throughout the workflow, the proposed framework positions agentic AI as a research-orchestration approach rather than an autonomous drug discovery or clinical decision-making system.

Key Words: *Agentic AI, Computational pharmacology, Molecular search, Evidence appraisal, Mechanistic reasoning, Candidate prioritization*

eIJPPR 2026; 16(2):123-135

HOW TO CITE THIS ARTICLE: Hassan L, Khalaf O, Jaber R. Agentic AI for Computational Pharmacology: Orchestrating Molecular Search, Evidence Appraisal, Mechanistic Reasoning, and Candidate Prioritization. Int J Pharm Phytopharmacol Res. 2026;16(2):123-35. https://doi.org/10.51847/5TFFxnr14x

INTRODUCTION

Agentic artificial intelligence represents an emerging computational paradigm in which models are organized to pursue multistep objectives through planning, tool selection, information retrieval, state tracking, and iterative interaction with human experts. In biomedical research, an agent may be configured to coordinate analytical models, databases, scientific literature, computational tools, and domain-specific review rather than merely generate an isolated textual response. This distinction is important because biomedical discovery involves linked questions

Corresponding author: Lina Hassan
Address: Department of AI for Dead Sea and Desert Plant Adaptogens, Faculty of Pharmacy, University of Jordan, Amman, Jordan.
E-mail: ✉ lina.hassan@ju.edu.jo
Received: 16 February 2026; **Revised:** 18 April 2026; **Accepted:** 21 April 2026

whose interpretation depends on experimental context, evidence quality, biological uncertainty, and the expertise of multiple disciplines. Collaborative biomedical agents may therefore support the navigation of complex research spaces, but they cannot independently determine whether a molecular association, mechanistic explanation, or therapeutic hypothesis is biologically true [1].

Agentic AI must also be distinguished from an ordinary chatbot or an unstructured sequence of prompts. A conversational language model generally responds to individual requests, whereas an agentic system may decompose an objective, construct a plan, call permitted tools, retain task-relevant state, evaluate intermediate outputs, and revise its actions. Multi-agent configurations extend this model by allocating distinct functions to planning, retrieval, analysis, verification, or specialist review agents. These capabilities may be useful in bioinformatics and biomedicine, although their value depends on reliable orchestration, appropriate tool access, controlled memory, transparent role definitions, and mechanisms for detecting unsupported reasoning [2]. Agentic terminology should consequently describe workflow organization and delegated computational functions, not imply human-equivalent scientific judgment or independent decision authority.

Computational pharmacology is a suitable domain for investigating this paradigm because pharmacological research rarely depends on a single computational prediction. A research question may require molecular-structure searches, bioactivity retrieval, target annotation, pathway interpretation, evaluation of disease context, assessment of pharmacokinetic and toxicity evidence, and comparison of competing candidate hypotheses. Machine-learning methods already contribute to target identification, molecular design, property prediction, biomarker development, and other stages of drug discovery and development, but their performance remains contingent on the relevance, representativeness, and quality of the underlying data [3]. An agentic architecture could help coordinate these heterogeneous tasks, provided that it preserves distinctions between experimental measurements, curated relationships, computational predictions, literature-derived statements, and agent-generated inferences.

The same coordination that makes agentic AI attractive also creates substantial risks. An agent may formulate an inappropriate query, retrieve an incomplete evidence set, select an unsuitable tool, misinterpret an assay, overlook contradictory findings, propagate a hallucinated statement, or rank a compound on the basis of weak surrogate endpoints. Faster molecular generation or screening does not necessarily produce better pharmacological decisions when computational objectives fail to represent exposure, efficacy, toxicity, disease heterogeneity, or translational constraints [4]. This article therefore proposes a human-governed framework for orchestrating molecular search, evidence appraisal, mechanistic reasoning, safety review, and candidate prioritization. Its purpose is not to describe an implemented agent network, but to define the functional stages, governance requirements, evidence boundaries, expert checkpoints, and validation needs that a responsible agentic AI workflow for computational pharmacology should contain.

Agentic AI in computational pharmacology

Within computational pharmacology, agentic AI can be defined as a governed computational architecture in which one or more specialized AI components coordinate a research objective through explicit task decomposition, planning, permitted tool use, evidence retrieval, structured reasoning, review, and feedback. This definition differs from a single-task predictive model that maps an input directly to an endpoint, a static retrieval system that returns documents without evaluating their relevance, and conventional workflow automation that executes a predetermined sequence without context-sensitive planning. Agentic bioinformatics architectures illustrate how complex biomedical questions may be separated into modular tasks and assigned to specialized analytical or retrieval functions [5]. In the proposed framework, however, specialization does not confer autonomous scientific authority: each agent performs a bounded computational function within an expert-defined workflow.

Task decomposition is the first operational requirement. A broad question such as whether a compound class merits further investigation must be reformulated into narrower questions concerning molecular identity, relevant targets, disease or phenotype context, pathway relationships, available bioactivity evidence, exposure constraints, potential toxicities, and experimental knowledge gaps. A planning agent may arrange these tasks, identify dependencies, and select approved tools, while specialist agents perform searches or analyses. Chemistry-focused tool-using systems demonstrate that language models can invoke external computational resources and coordinate multistep chemical tasks, but they also show that tool augmentation does not eliminate evaluation errors, ambiguous outputs, or the need for safety controls [6]. The planning layer should therefore specify permitted actions, required evidence types, stopping conditions, and mandatory review points before execution begins.

Autonomy should be treated as a graded property rather than a binary label. A tool-using system may autonomously select among low-risk database queries while requiring approval before code execution, external

communication, laboratory control, or any action that could affect physical experiments. Systems developed for automated chemical research show that search, code generation, procedural planning, and experimental tools can be connected within a common architecture [7]. Nevertheless, the ability to formulate or execute a chemical procedure does not establish the pharmacological validity, safety, or translational value of its output. A responsible architecture must differentiate planning authority from execution authority, record every tool call, restrict credentials through least-privilege access, and permit experts to interrupt, revise, or terminate the workflow.

The intended role of agentic AI is consequently one of structured research support. Retrieval agents may organize literature and database evidence; molecular search agents may identify structurally or functionally relevant records; appraisal agents may assess source quality and evidential fit; reasoning agents may construct explicit mechanistic hypotheses; and prioritization agents may compare hypotheses using transparent criteria. Recent agentic drug-discovery scholarship suggests that modular coordination could make computational workflows more accessible and adaptable, while simultaneously emphasizing unresolved challenges in validation, reliability, and expert governance [8]. Human pharmacologists, medicinal chemists, toxicologists, computational scientists, and other reviewers must retain responsibility for interpreting evidence, determining whether a candidate hypothesis should progress, and deciding which experiments are necessary. Agentic AI should support research decision-making rather than autonomously authorize pharmacological or clinical action.

Molecular search and evidence appraisal

Molecular search begins with a clearly specified pharmacological question and an auditable query plan. The workflow should define the candidate class, structural representation, molecular identifiers, similarity or substructure criteria, target context, disease or phenotype context, and any property filters that are scientifically justified. A molecular search agent may then query permitted chemical, bioactivity, structural, or biomolecular resources and return traceable records for further evaluation. Modular drug-discovery agents demonstrate that retrieval, molecular generation, property prediction, structural analysis, and iterative refinement can be connected as separate computational tasks [9]. These capabilities support workflow organization, but each returned structure, predicted property, or database association must remain a provisional search output until its provenance, chemical validity, assay context, and relevance have been reviewed. Database interfaces can improve the precision of biomedical information access, although incorrect API selection, malformed tool calls, entity ambiguity, and unsupported interpretation remain possible [10].

Literature retrieval must be separated from evidence acceptance. A retrieval agent can search peer-reviewed literature using compound names, standardized identifiers, target synonyms, pathway terms, disease concepts, assay types, and citation relationships. It may also retrieve full metadata and link statements to their source documents. Biomedical language-model platforms show how modular retrieval, knowledge integration, model chaining, and benchmarking can be combined within controlled research applications [11]. Nevertheless, a retrieved article is not automatically relevant to a specific claim, and a source-linked answer is not necessarily correct. Retrieval-augmented systems may improve factual completeness and source grounding in evidence-intensive biomedical tasks, but their performance remains dependent on retrieval quality, source coverage, question formulation, and expert evaluation [12]. The appraisal stage must therefore examine study type, methods, experimental system, directness, sample context, endpoint validity, and the degree to which each source supports the exact proposition under consideration.

Structured knowledge resources can complement literature retrieval by organizing relationships among compounds, drugs, genes, proteins, diseases, phenotypes, and other biomedical entities. Integrated knowledge graphs may allow a search agent to construct contextual subgraphs connecting a candidate molecule with potential targets, disease associations, molecular functions, and therapeutic relationships [13]. Such graphs can improve discoverability and expose otherwise dispersed relationships, but their edges may represent different levels of evidence, including curated observations, computational predictions, literature extractions, or inherited database assertions. The evidence-appraisal agent must preserve these distinctions rather than collapse them into a uniform interaction network. Mechanistic statements extracted from text and curated databases can similarly be assembled at scale while retaining source-linked evidence, enabling provenance-aware inspection of proposed causal or regulatory relationships [14]. These assembled statements remain representations of reported evidence and cannot by themselves confirm a biological mechanism.

Evidence appraisal should produce a structured evidence profile rather than a single opaque quality score. For each claim or candidate hypothesis, the profile should identify source type, evidential directness, experimental context, reproducibility, conflicts, missing information, and dependence on computational prediction. Provenance records should connect every extracted datum and generated statement to its source, query, database version,

and processing step. Citation traceability should confirm semantic correspondence between a manuscript claim and its supporting reference, while conflict detection should distinguish genuine disagreement from differences in assay conditions, biological systems, doses, populations, or endpoints. Uncertainty should be stated in terms of missing data, source limitations, model applicability, and unresolved contradictions, and the audit trail should preserve searches, tool calls, transformations, reviewer interventions, and version changes. **Table 1** summarises molecular search and evidence appraisal components required for agentic AI workflows in computational pharmacology.

Table 1. Molecular Search and Evidence Appraisal Components for Agentic AI in Computational Pharmacology: Task Decomposition, Query Planning, Literature Retrieval, Molecular Database Search, Target and Pathway Retrieval, Source Quality, Evidence Provenance, Conflict Detection, Uncertainty, and Auditability

Workflow component	Core question	Agentic function	Input evidence	Potential output	Reliability concern	Expert review need	Validation boundary	Framework role
Task decomposition	Which computational and pharmacological subtasks are required to address the research question?	Divide the question into molecular search, retrieval, appraisal, mechanism, safety, and prioritization tasks.	User-defined objective, disease or phenotype context, candidate scope, intended use.	Structured task graph with dependencies and stopping points.	Important evidence domains may be omitted or incorrectly sequenced.	A pharmacologist confirms the scope, exclusions, and required domains.	A task graph is a workflow proposal and does not validate a hypothesis.	Establishes bounded responsibilities for each agent.
Query planning	Which entities, synonyms, identifiers, filters, and search strategies are required?	Generate database-specific and literature-specific search plans.	Compound names, standardized identifiers, target names, pathways, diseases, endpoints.	Versioned query protocol and search log.	Ambiguous terminology, incomplete synonyms, and overly restrictive filters.	An information specialist or domain expert reviews high-impact queries.	Search completeness cannot be assumed from successful query execution.	Converts the research task into reproducible retrieval operations.
Molecular search	Which molecular records satisfy the initial structural or property constraints?	Search chemical spaces using identifiers, similarity, substructure, or approved filters.	Molecular structures, fingerprints, identifiers, physicochemical criteria.	Traceable molecular hit set or candidate-hypothesis set.	Stereochemical errors, duplicates, salts, invalid structures, and search bias.	A medicinal chemist reviews structural validity and chemical relevance.	Molecular hits are not validated candidates.	Generates molecular hypotheses for subsequent appraisal.
Literature retrieval	Which peer-reviewed studies address the compound, target, pathway, phenotype,	Retrieve and organize relevant journal literature.	Bibliographic queries, entity names, citation networks, study-type filters.	Source corpus linked to entities and proposed claims.	Publication bias, retrieval omissions, irrelevant sources, or superseded findings.	Experts assess relevance, methods, and contextual fit.	Retrieval does not establish evidence quality or causal validity.	Supplies literature evidence to the appraisal layer.

	or safety question?							
Molecular database search	What structure, assay, bioactivity, and property records are available?	Query approved chemical and bioactivity resources through governed tools.	Standardized molecular identifiers and database-specific fields.	Structure records, assay links, activity records, and provenance metadata.	Assay heterogeneity, inconsistent units, duplicate records, and outdated entries.	Chemoinformatics review of normalization and assay comparability.	Database values require source verification and applicability assessment.	Provides machine-readable molecular evidence.
Target database search	What evidence connects a molecule with a biological target?	Retrieve target annotations, interaction records, and assay-linked evidence.	Compound and target identifiers, assay metadata, curated interaction records.	Drug–target evidence profile categorized by evidence type.	Predicted, indirect, and experimentally measured interactions may be conflated.	A pharmacologist distinguishes binding, functional, indirect, and predicted evidence.	A database association does not establish target engagement in the intended context.	Grounds target-context and mechanistic reasoning.
Pathway knowledge retrieval	Which pathways connect the target, phenotype, and disease context?	Retrieve curated signaling, regulatory, and pathway relationships.	Target identifiers, cell or tissue context, disease concepts, pathway resources.	Contextual pathway subgraph with source provenance.	Generic pathway knowledge may be misapplied to a specific biological context.	A systems pharmacologist reviews context, directionality, and feedback.	Pathway membership does not prove causal therapeutic modulation.	Supplies structured inputs for pathway reasoning.
Source quality assessment	Is each source methodologically and contextually suitable for the proposed claim?	Classify source type, methods, directness, and relevance.	Full-text methods, study design, experimental system, endpoints, metadata.	Source-quality and claim-fit profile.	Journal prestige or citation count may be mistaken for evidential directness.	Human review confirms whether the source supports the exact claim.	Quality classification cannot substitute for critical reading.	Prevents weakly matched or unsupported citation use.
Evidence provenance tracking	Where did each datum, relationship, and generated statement originate?	Attach source, query, database version, extraction method, and transformation history.	Article identifiers, record identifiers, query logs, extraction records.	Provenance graph connecting outputs to their origins.	Merging or transformation may obscure the original evidence lineage.	Reviewers inspect provenance for high-impact claims and decisions.	Provenance demonstrates traceability, not scientific truth.	Enables reproducibility and accountability across agents.
Citation traceability	Does every evidence-based statement map to an appropriate	Link claims to specific source records and	Candidate statements, extracted findings, bibliographic metadata.	Claim–citation map and unsupported-claim warnings.	Keyword similarity may be mistaken for substantiv	Human verification of semantic correspondence is required.	A citation link does not guarantee correct	Supports citation-safe evidence communication.

	supporting source?	supporting evidence.			e evidential support.		interpretation.	
Data conflict detection	Do sources disagree about activity, target relationships, mechanism, exposure, or safety?	Compare findings by direction, assay, context, and evidence type.	Normalized extracted results and contextual metadata.	Conflict report with possible contextual explanations.	Incomparable experiments may be incorrectly classified as contradictory.	Experts determine whether findings conflict or address different conditions.	The agent must not resolve substantive disputes without justified review.	Prevents selective evidence use and unsupported reconciliation.
Evidence grading	How direct, reproducible, and contextually relevant is the available evidence?	Assign transparent evidence descriptors across predefined dimensions.	Study design, replication, directness, biological context, source quality.	Multidimensional evidence profile.	Numerical aggregation may create false precision or hide weak dimensions.	Experts approve grading criteria and interpret exceptions.	Evidence grades facilitate comparison but do not certify validity.	Provides transparent inputs to reasoning and prioritization.
Uncertainty communication	What is unknown, extrapolated, missing, inconsistent, or outside model applicability?	Generate evidence-linked uncertainty statements and abstention flags.	Data gaps, conflicts, prediction limits, source limitations, applicability information.	Uncertainty register and requests for additional evidence.	Unjustified confidence categories may create false reassurance.	Experts confirm that uncertainty is communicated proportionately.	Uncertainty estimation does not replace empirical validation.	Constrains interpretation of all downstream outputs.
Audit trail	Can every query, tool call, transformation, decision, and override be reconstructed?	Record inputs, outputs, permissions, versions, timestamps, and reviewer actions.	Logs from agents, tools, databases, and expert-review stages.	Reconstructable and versioned workflow record.	Missing logs, mutable records, or undocumented tool changes.	Governance review confirms completeness before outputs are released.	Auditability supports accountability but does not validate scientific conclusions.	Provides cross-cutting governance for the full workflow.

Mechanistic reasoning

Mechanistic reasoning in computational pharmacology should begin by asking whether a proposed target is relevant within the specified disease, phenotype, tissue, cell state, and treatment context. Heterogeneous biomedical networks can connect compounds, targets, genes, pathways, diseases, and phenotypes to support the construction of evidence-linked repurposing hypotheses, but such network-derived relationships remain starting points for investigation rather than confirmation of therapeutic mechanism [15]. A mechanistic reasoning agent should therefore distinguish direct experimental evidence from curated associations, computational predictions, and inferred network paths. It should also retrieve signaling relationships with sufficient cellular and molecular context, because the pharmacological meaning of target modulation depends on pathway directionality, upstream and downstream regulation, intercellular communication, feedback mechanisms, and the biological system in which those relationships were observed [16].

The reasoning process should combine target-context interpretation with disease-specific and therapeutic evidence without collapsing these sources into a single claim of validity. Graph-based repurposing models can integrate disease and drug relationships to produce contextualized candidate hypotheses and explanatory

evidence, yet their outputs remain predictions whose relevance must be assessed by specialists [17]. Safety-related relationships should be incorporated at the same stage rather than added only after a candidate has been ranked. Computational modeling of polypharmacy side effects illustrates the importance of considering drug–drug, drug–protein, and relation-specific adverse-effect patterns when evaluating candidate hypotheses [18]. A mechanistic reasoning agent should consequently flag potential interaction pathways, competing targets, and safety-relevant network connections while clearly separating observed effects from predicted associations.

ADMET and toxicity information provides another essential constraint on mechanistic plausibility. A theoretically relevant target relationship may have limited pharmacological value when the compound cannot plausibly reach the required tissue, is rapidly metabolized, interacts with critical transporters, or carries unresolved toxicity liabilities. Computational ADMET resources can assist early evidence organization and comparative screening, but their predictions cannot establish absorption, exposure, metabolism, excretion, tolerability, or safety in the intended biological setting [19]. The proposed reasoning workflow should record whether each pharmacokinetic or toxicity statement is experimentally measured, literature-reported, database-curated, structurally inferred, or model-predicted. Missing evidence should be represented explicitly, and the absence of a reported liability must never be interpreted as evidence that the liability is absent.

The output of mechanistic reasoning should be a structured plausibility assessment containing supporting evidence, contradictory findings, alternative explanations, contextual assumptions, uncertainty sources, and proposed validation needs. Explainable molecular AI may help identify structural features or input relationships associated with a prediction, but an explanation of model behavior does not establish a causal biological mechanism [20]. Contradiction detection should compare assay systems, doses, exposure conditions, disease models, and evidence types before classifying findings as genuinely inconsistent. Expert pharmacologists should review whether the proposed compound–target–pathway–phenotype chain is coherent, whether crucial causal links are missing, and whether the planned experiments can test those links. Agentic reasoning may organize mechanistic hypotheses and expose evidential gaps, but external experimental evidence is required to establish target engagement, pathway modulation, mechanism, safety, or therapeutic effect.

Candidate prioritization logic

Candidate prioritization should be treated as a transparent multi-objective comparison rather than a single-score selection exercise. Relevant dimensions may include bioactivity evidence, target relevance, pathway plausibility, disease-context fit, physicochemical properties, predicted or measured ADMET characteristics, toxicity evidence, interaction risk, chemical feasibility, novelty, evidence quality, and uncertainty. AI-enabled drug design is most useful when computational methods augment medicinal-chemistry and pharmacology judgment across these competing objectives rather than replacing expert deliberation [21]. The proposed candidate-prioritization agent should therefore expose the evidence and assumptions associated with every criterion, identify missing dimensions, and show how different weighting choices alter the resulting order.

Molecular generation and search outputs should enter the prioritization stage only as candidate hypotheses. Generative models can produce target-focused molecular libraries and support the exploration of chemical space, but generated structures do not become validated compounds merely because they satisfy computational constraints [22]. Likewise, a high model score or rank does not establish biological activity, selectivity, exposure, safety, or efficacy. The value of computational prioritization depends on what happens after ranking: studies that connect model-based screening with orthogonal laboratory and biological testing demonstrate that experimental confirmation is necessary before a computationally prioritized molecule can be interpreted as pharmacologically meaningful [23]. The framework should therefore prevent the labels “validated candidate,” “safe candidate,” or “effective candidate” from being assigned on the basis of ranking alone.

A multi-criteria prioritization agent may compare potency-related evidence with selectivity, physicochemical properties, exposure requirements, toxicity concerns, synthesis feasibility, and evidential strength. Explainable multiparameter optimization can help make the contribution of individual criteria visible, allowing experts to inspect why one candidate is favored over another [24]. Nevertheless, composite scores can conceal serious weaknesses when strong performance on several dimensions offsets a critical liability. The framework should preserve non-compensatory stopping rules for severe toxicity concerns, unsupported target assumptions, structurally invalid records, or missing evidence essential to the research question. It should also report rank sensitivity when plausible changes in weights, thresholds, or evidence interpretation substantially alter the prioritization.

Alternative ranking views should be retained when no single candidate dominates across all objectives. Multi-

objective drug-design methods include weighted aggregation, conditional optimization, and Pareto-based comparison, each of which encodes different assumptions about trade-offs and acceptable compromise [25]. The prioritization agent should therefore provide an evidence-linked candidate profile, a transparent comparison of trade-offs, and an uncertainty statement rather than an unexplained ordered list. Expert pharmacology, medicinal-chemistry, toxicology, and experimental review should determine whether any hypothesis warrants further testing. Rankings support research prioritization only; they do not authorize candidate progression, establish therapeutic value, or support clinical claims.

Proposed agentic AI framework

The proposed framework begins with human definition of the research question, intended use, disease or phenotype context, molecular scope, target hypothesis, safety concerns, and candidate-prioritization objective. A task-decomposition stage converts this specification into bounded questions concerning molecular identity, literature evidence, target relationships, pathway context, ADMET, toxicity, and validation needs. An agent-planning layer then determines the sequence of permitted searches, analyses, appraisal steps, and review gates. Because model-supported decisions can carry substantial uncertainty, the framework requires uncertainty to be recorded throughout the workflow rather than added only to the final output [26]. Planning agents must be able to abstain, request clarification from an expert, or halt a task when essential inputs are missing or when the available evidence falls outside the approved scope.

Tool-use governance surrounds the planning, molecular-search, and retrieval stages. Each agent receives access only to approved tools, databases, functions, and credentials necessary for its assigned task. Search queries, database versions, inputs, outputs, failures, and transformations are recorded in an audit trail, and high-risk actions require explicit human authorization. Responsible machine-learning practice requires clear specification of context, assumptions, validation procedures, reproducibility controls, and mechanisms for detecting potential harm [27]. In the proposed framework, molecular search and evidence retrieval are therefore separated from evidence acceptance: retrieval agents collect potentially relevant records, while appraisal agents evaluate source quality, provenance, citation fit, conflicts, and uncertainty before evidence can influence reasoning or prioritization.

Appraised evidence proceeds to mechanistic reasoning and safety review. The mechanistic reasoning agent organizes target-context evidence, pathway relationships, drug–target interactions, phenotype associations, competing explanations, and evidential gaps. In parallel, the safety-review agent examines ADMET evidence, toxicity findings, interaction risks, exposure limitations, and unresolved uncertainties. Only evidence that retains its source type and provenance may enter candidate prioritization. The prioritization agent compares candidate hypotheses across transparent project criteria and provides trade-off profiles rather than autonomous progression decisions. Technical performance or computational coherence alone cannot establish real-world utility, particularly when workflow integration, external validation, and stakeholder interpretation have not been examined [28]. Mandatory expert review therefore separates computational prioritization from validation planning.

Auditability and explanation operate across all framework stages. Every claim, intermediate representation, tool call, evidence transformation, uncertainty flag, reviewer intervention, and version change should be reconstructable. Explanations must be adapted to the needs of pharmacologists, medicinal chemists, toxicologists, computational scientists, and other responsible reviewers, because the same model output may require different forms of justification for different decisions. Explainability should not be treated as automatic evidence of correctness, causality, fairness, or trustworthiness [29]. The framework therefore requires reviewers to inspect the underlying evidence and assumptions rather than accept a persuasive narrative or visual explanation as proof.

Feedback updating closes the workflow. Expert corrections, newly retrieved evidence, negative findings, validation results, and database updates should return to the relevant search, appraisal, reasoning, or prioritization stage through a version-controlled process that preserves the earlier record. The framework prevents silent replacement of previous evidence and requires documentation of why a conclusion, uncertainty statement, or candidate order changed. It also enforces a no-clinical-claim boundary and a no-autonomous-decision boundary: outputs may organize research hypotheses and validation plans but may not recommend patient care, establish clinical utility, or authorize candidate advancement. Any future transition from research support toward clinical decision support would require staged evaluation, human-factors assessment, safety reporting, and transparent documentation of limitations [30]. **Table 2** presents the proposed agentic AI framework for computational pharmacology.

Table 2. Proposed Agentic AI Framework for Computational Pharmacology: Task Decomposition, Tool-Use Governance, Molecular Search, Evidence Appraisal, Mechanistic Reasoning, Safety Review, Candidate Prioritization, Expert Oversight, Validation Planning, Feedback, and Decision Boundaries

Framework stage	Core purpose	Required input	Agentic function	Potential output	Governance requirement	Failure risk	Feedback mechanism	Decision boundary
User research question definition	Establish the scientific objective, intended use, scope, and exclusions.	Research question, disease or phenotype context, candidate class, target hypothesis, safety concerns.	Convert the request into a structured research specification.	Expert-approved scope statement.	A named expert must confirm the intended use and required evidence domains.	Ambiguity may propagate irrelevant searches or unsupported conclusions.	Unresolved terms or assumptions are returned to the initiating expert.	Restricted to research support; no patient-specific decision.
Task decomposition	Divide the question into auditable computational and pharmacological subtasks.	Approved scope statement.	Create linked tasks for search, retrieval, appraisal, mechanism, safety, prioritization, and validation planning.	Versioned task graph with dependencies and stopping rules.	Mandatory review stages must be defined before execution.	Safety, context, or validation tasks may be omitted.	Experts may add, remove, or reorder tasks.	Decomposition does not authorize autonomous high-risk actions.
Agent planning	Determine the sequence of agents, tools, and review checkpoints.	Task graph, tool permissions, evidence requirements, resource constraints.	Produce a reviewable execution plan.	Agent plan with assigned roles and escalation rules.	Plan approval, resource limits, and predefined stopping conditions.	Circular planning, uncontrolled tool use, or inappropriate delegation.	Failed or conflicting steps trigger documented replanning.	Planning agents cannot alter governance rules or intended use.
Tool-use governance	Restrict tool access and preserve control over data and actions.	Approved tool registry, permissions, credentials, risk classification.	Authorize, block, or escalate proposed tool calls.	Permitted tool execution or documented refusal.	Least-privilege access, logging, sandboxing, and human approval for high-risk actions.	Data leakage, unsafe execution, incorrect external action, or undocumented tool changes.	Tool failures and violations return to governance and planning review.	No unrestricted laboratory, clinical, or external-system control.
Molecular search	Identify molecular records relevant to predefined structural or pharmacological constraints.	Structures, identifiers, similarity rules, substructure criteria, approved property filters.	Search permitted chemical and bioactivity resources.	Traceable molecular hit set or hypothesis set.	Standardization, database-version tracking, and chemical-validity checks.	Invalid structures, duplicates, stereochemical errors, and search bias.	Search criteria are revised following appraisal or expert review.	Search hits are not validated candidates.
Evidence retrieval	Collect literature and structured evidence relevant to each candidate hypothesis.	Approved queries, compounds, targets, pathways, diseases, safety endpoints.	Retrieve peer-reviewed literature and permitted database records.	Source corpus and retrieval log.	Source constraints, provenance retention, and retraction or update checks.	Missing, outdated, irrelevant, or non-comparable evidence.	Appraisal may request broader, narrower, or alternative searches.	Retrieved information is not accepted evidence automatically.

Evidence appraisal	Assess source quality, evidential directness, claim fit, conflicts, and uncertainty.	Retrieved studies, records, methods, metadata, and extracted findings.	Produce evidence profiles and claim–source mappings.	Appraised evidence package and conflict report.	Human verification of high-impact extractions and citation correspondence.	Hallucinated extraction, weak source fit, or false evidence equivalence.	Conflicts or gaps reopen retrieval and expert review.	Appraisal cannot certify biological truth.
Mechanistic reasoning	Organize evidence into target, pathway, phenotype, and disease-context hypotheses.	Appraised interaction, pathway, target, phenotype, and contextual evidence.	Construct plausible causal chains, alternatives, and missing-link maps.	Mechanistic plausibility profile with uncertainties.	Observed, predicted, curated, and inferred relationships must remain distinct.	Correlation, proximity, or model explanation may be mistaken for mechanism.	Contradictions or gaps return to retrieval and appraisal.	Does not establish mechanism of action.
Safety review	Identify pharmacokinetic, toxicity, interaction, and exposure-related concerns.	ADMET evidence, toxicity studies, structural alerts, interaction records, uncertainty information.	Aggregate safety evidence and apply stopping or escalation flags.	Safety-risk register and evidence-gap list.	Toxicology oversight and non-overridable severe-risk escalation.	Missing safety data may be interpreted as evidence of safety.	Additional evidence requests or candidate hold.	The agent cannot declare a candidate safe.
Candidate prioritization	Compare candidate hypotheses using transparent and reviewable criteria.	Mechanistic evidence, bioactivity, ADMET, toxicity, feasibility, evidence quality, uncertainty.	Generate multi-criteria profiles, trade-off views, and provisional ordering.	Evidence-linked candidate-prioritization hypotheses.	Criteria, weights, exclusions, and sensitivity analyses must be visible.	Hidden weighting, rank instability, false precision, or automation bias.	Expert changes produce a versioned reprioritization.	Ranking cannot authorize candidate progression.
Expert oversight	Provide accountable scientific interpretation and adjudication.	Full evidence package, audit log, conflicts, uncertainties, and provisional rankings.	Present structured evidence and record reviewer decisions.	Approval for further investigation, revision, rejection, or evidence request.	Named pharmacology, medicinal-chemistry, toxicology, and computational reviewers as appropriate.	Rubber-stamping or excessive trust in coherent agent outputs.	Reviewer comments trigger targeted reanalysis.	Humans retain final research decision authority.
Validation planning	Define experiments needed to test the most consequential assumptions and risks.	Expert-reviewed hypotheses, mechanistic gaps, safety concerns, uncertainty register.	Match hypotheses with orthogonal assays, controls, endpoints, and stopping criteria.	Proposed staged validation plan.	Laboratory, statistical, feasibility, ethical, and domain review.	Proposed tests may fail to address the causal or safety question.	Experimental findings return to the appropriate evidence stages.	A validation plan must not be represented as completed validation.
Audit trail	Preserve a reconstructable history of the entire workflow.	Plans, queries, tool calls, outputs, transformations, permissions,	Maintain immutable or controlled versioned records.	Complete audit package.	Access control, retention rules, versioning, and change documentation.	Missing lineage, untraceable edits, or incomplete logs.	Audit findings may reopen any earlier stage.	Auditability supports accountability but does not validate conclusions.

		reviews, and overrides.						
Feedback updating	Incorporate new evidence and expert corrections without erasing prior reasoning.	Validation findings, new literature, database updates, expert feedback, negative results.	Update affected evidence nodes and rerun justified tasks.	Versioned evidence profiles and revised prioritization.	Change control and comparison with previous versions.	Confirmation bias, silent ranking changes, or model and prompt drift.	Before-and-after review identifies why outputs changed.	Updated outputs remain hypotheses until reappraised and reviewed.
No autonomous-decision boundary	Prevent agent outputs from becoming unsupervised pharmacological or clinical decisions.	Workflow state, permissions, intended-use rules, reviewer status.	Block unauthorized candidate approval, clinical language, or progression actions.	Escalation notice or workflow halt.	Mandatory human authorization and institutional accountability.	Recommendations may become de facto decisions through automatic acceptance.	Governance review is required before restart.	No autonomous candidate approval, therapeutic recommendation, or clinical decision.

Figure 1 illustrates the proposed agentic AI framework for computational pharmacology, connecting task decomposition, molecular search, evidence appraisal, mechanistic reasoning, candidate prioritization, expert oversight, validation planning, and feedback updating.

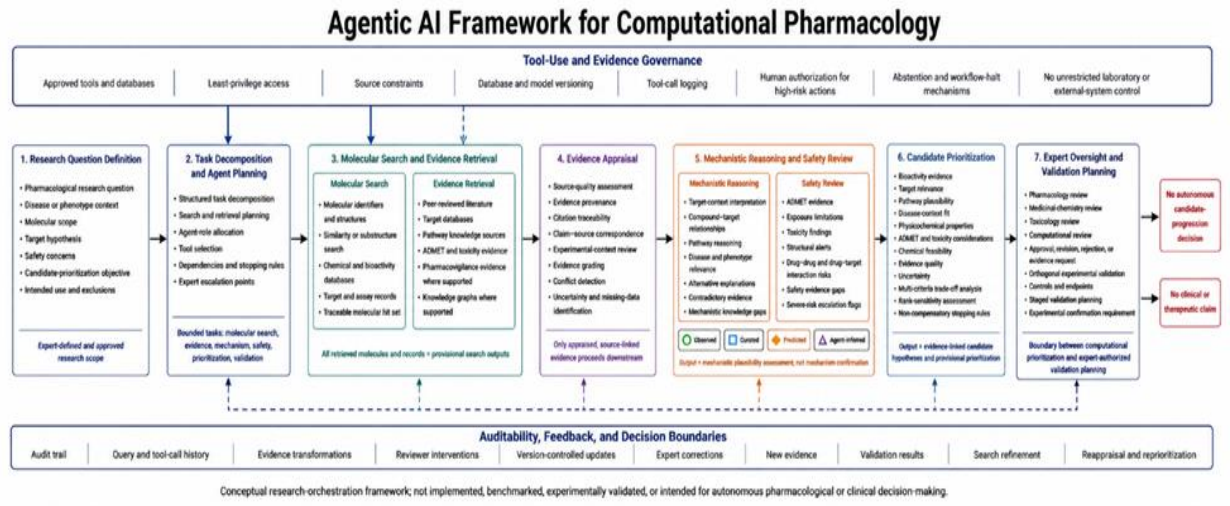

Figure 1. Agentic AI Framework for Computational Pharmacology

Alt-text description

A conceptual agentic AI framework showing a research question decomposed into molecular search, evidence appraisal, mechanistic reasoning, safety review, candidate prioritization, expert oversight, validation planning, audit trail, feedback loops, and decision boundaries.

CONCLUSION

This article proposes a human-governed agentic AI framework for computational pharmacology that coordinates research question definition, task decomposition, controlled tool use, molecular search, literature and database retrieval, evidence appraisal, mechanistic reasoning, safety review, candidate prioritization, expert oversight, validation planning, auditability, and feedback updating. Its central contribution is the separation of retrieval from evidence acceptance, mechanistic plausibility from mechanism validation, and computational ranking from pharmacological decision-making. Agentic AI may support the organization of complex research workflows and help experts identify evidence gaps, contradictions, safety concerns, and testable hypotheses, but it cannot independently establish biological truth, candidate validity, safety, efficacy, translational readiness, or clinical utility. Source

ISSN (Online) 2249-6084 (Print) 2250-1029

grounding, provenance, citation traceability, uncertainty communication, transparent prioritization criteria, expert review, and explicit stopping rules must therefore be built into every stage. Under these conditions, agentic AI can function as a structured research-orchestration and prioritization approach while preserving the authority of pharmacologists, medicinal chemists, toxicologists, experimental scientists, clinicians, and other accountable reviewers.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Gao S, Fang A, Huang Y, Giunchiglia V, Noori A, Schwarz JR, et al. Empowering biomedical discovery with AI agents. Cell. 2024;187(22):6125-51. doi:10.1016/j.cell.2024.09.022

[2] Yang T, Xiao Y, Bao Z, Hao J, Peng J. The rise and potential opportunities of large language model agents in bioinformatics and biomedicine. Brief Bioinform. 2025;26(6). doi:10.1093/bib/bbaf601

[3] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-77. doi:10.1038/s41573-019-0024-5

[4] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discov Today. 2021;26(2):511-24. doi:10.1016/j.drudis.2020.12.009

[5] Zhou J, Jiang J, Han Z, Wang Z, Gao X. Streamline automated biomedical discoveries with agentic bioinformatics. Brief Bioinform. 2025;26(5). doi:10.1093/bib/bbaf505

[6] Bran AM, Cox S, Schilter O, Baldassari C, White AD, Schwaller P. Augmenting large language models with chemistry tools. Nat Mach Intell. 2024;6(5):525-35. doi:10.1038/s42256-024-00832-8

[7] 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

[8] He J, Lai H, Saigiridharan L, Ghiandoni GM, Jenei K, Gokalp U, et al. Democratising real-world drug discovery through agentic AI. Drug Discov Today. 2026;31(2):104605. doi:10.1016/j.drudis.2026.104605

[9] Ock J, Meda RS, Badrinarayanan S, Aluru NS, Chandrasekhar A, Barati Farimani A. Large language model agent for modular task execution in drug discovery. J Chem Inf Model. 2026;66(4):2055-68. doi:10.1021/acs.jcim.5c02454

[10] Jin Q, Yang Y, Chen Q, Lu Z. GeneGPT: Augmenting large language models with domain tools for improved access to biomedical information. Bioinformatics. 2024;40(2). doi:10.1093/bioinformatics/btae075

[11] Lobentanzer S, Feng S, Bruderer N, Maier A, BioChatter Consortium, Wang C, et al. A platform for the biomedical application of large language models. Nat Biotechnol. 2025;43(2):166-9. doi:10.1038/s41587-024-02534-3

[12] Zakka C, Shad R, Chaurasia A, Dalal AR, Kim JL, Moor M, et al. Almanac—retrieval-augmented language models for clinical medicine. NEJM AI. 2024;1(2). doi:10.1056/AIoa2300068

[13] Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data. 2023;10(1):67. doi:10.1038/s41597-023-01960-3

[14] Bachman JA, Gyori BM, Sorger PK. Automated assembly of molecular mechanisms at scale from text mining and curated databases. Mol Syst Biol. 2023;19(5). doi:10.15252/msb.202211325

[15] Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Elife. 2017;6. doi:10.7554/eLife.26726

[16] Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Klein M, Ivanova O, et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol Syst Biol. 2021;17(3). doi:10.15252/msb.20209923

[17] Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. Nat Med. 2024;30(12):3601-13. doi:10.1038/s41591-024-03233-x

[18] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13). doi:10.1093/bioinformatics/bty294

[19] Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z, et al. admetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties. Bioinformatics. 2019;35(6):1067-9. doi:10.1093/bioinformatics/bty707

[20] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat

Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4

[21] Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-64. doi:10.1038/s41573-019-0050-3

[22] Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018;4(1):120-31. doi:10.1021/acscentsci.7b00512

[23] Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688-702.e13. doi:10.1016/j.cell.2020.01.021

[24] Bung N, Krishnan SR, Roy A. An in silico explainable multiparameter optimization approach for de novo drug design against proteins from the central nervous system. J Chem Inf Model. 2022;62(11):2685-95. doi:10.1021/acs.jcim.2c00462

[25] Luukkonen S, van den Maagdenberg HW, Emmerich MTM, van Westen GJP. Artificial intelligence in multi-objective drug design. Curr Opin Struct Biol. 2023;79:102537. doi:10.1016/j.sbi.2023.102537

[26] Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nat Mach Intell. 2019;1(1):20-3. doi:10.1038/s42256-018-0004-1

[27] Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6

[28] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2

[29] Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q Consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi:10.1186/s12911-020-01332-6

[30] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-33. doi:10.1038/s41591-022-01772-9