Human-in-the-Loop Computational Phytopharmacology: Expert Curation, Model Oversight, Mechanistic Reasoning, and Translational Decision Confidence

Linus Chen[1]*, Yuki Sato[2], Olivia Harris[1]

[1]Department of Artificial Intelligence for Polyketide and Alkaloid Discovery, School of Pharmacy, University of California, San Francisco, San Francisco, United States.
[2]Department of Machine Learning for Biosynthetic Gene Clusters, Graduate School of Pharmacy, University of Tokyo, Tokyo, Japan.

ABSTRACT

Computational phytopharmacology increasingly uses artificial intelligence, natural product informatics, target prediction, pathway analysis, safety modeling, and evidence integration to support discovery-stage decisions. However, phytopharmacological data are often chemically complex, biologically heterogeneous, and dependent on uncertain botanical identity, chemical annotation, assay context, target evidence, pathway inference, and safety interpretation. This article proposes an original human-in-the-loop framework for computational phytopharmacology in which expert curation, model oversight, mechanistic reasoning, uncertainty interpretation, translational review, and decision-confidence assessment are embedded across the discovery workflow. The framework distinguishes expert curation from data annotation, model oversight from model selection, explainability review from causal interpretation, decision confidence from clinical readiness, and translational review from regulatory endorsement. It positions human expertise before, during, and after computational inference through checkpoints for botanical identity, chemical annotation, bioactivity evidence, assay relevance, training-data suitability, model-output interpretation, uncertainty communication, bias detection, biological plausibility, exposure plausibility, safety review, herb–drug interaction review, evidence grading, validation planning, and decision-boundary control. The main contribution is a structured framework that treats computational outputs as research-prioritisation evidence requiring transparent documentation, expert disagreement recording, validation planning, audit trails, and feedback loops for data and model refinement. The framework is conceptual and does not represent an implemented, clinically validated, or regulatory-approved decision-support system.

Key Words: *Human-in-the-loop, Computational phytopharmacology, Expert curation, Model oversight, Mechanistic reasoning, Decision confidence*

eIJPPR 2025; 15(6):1-10

HOW TO CITE THIS ARTICLE: Chen L, Sato Y, Harris O. Human-in-the-Loop Computational Phytopharmacology: Expert Curation, Model Oversight, Mechanistic Reasoning, and Translational Decision Confidence. Int J Pharm Phytopharmacol Res. 2025;15(6):1-10. https://doi.org/10.51847/QfrhT9l78x

INTRODUCTION

Computational phytopharmacology is increasingly shaped by machine learning, natural product informatics, molecular prediction, target inference, pathway analysis, safety modelling, and translational prioritisation. In drug discovery more broadly, machine learning can support decision-making when the biological question, data structure, endpoint definition, and modelling task are suitable for the intended use [1].

For phytopharmacology, this condition is especially important because plant-derived materials may involve

Corresponding author: Linus Chen
Address: Department of Artificial Intelligence for Polyketide and Alkaloid Discovery, School of Pharmacy, University of California, San Francisco, San Francisco, United States.
E-mail: ✉ linus.chen@ucsf.edu
Received: 19 July 2025; **Revised:** 28 October 2025; **Accepted:** 02 November 2025

complex mixtures, uncertain compound identity, variable extract composition, and bioactivity evidence generated across heterogeneous experimental systems.

The expanding use of artificial intelligence in drug discovery has also created a need for realistic interpretation of computational outputs. Model predictions may help prioritise hypotheses, compounds, targets, pathways, or safety concerns, but they should not be treated as autonomous evidence of pharmacological validity. Critical analyses of AI in drug discovery emphasise that computational impact depends on data quality, appropriate validation, domain constraints, and careful interpretation of what models can and cannot support [2]. This caution is central to phytopharmacology, where uncertain chemical annotations or weak assay relevance can propagate into misleading mechanistic narratives.

Computational prioritisation can be powerful when linked to experimental follow-up. Deep-learning approaches have shown that models may identify candidates for biological testing, but the evidentiary value of such predictions depends on downstream validation rather than prediction alone [3]. In computational phytopharmacology, this distinction is necessary because predicted bioactivity, target engagement, pathway modulation, or safety liability remains a research hypothesis until supported by fit-for-purpose chemical, pharmacological, toxicological, or translational evidence.

Artificial intelligence methods can support multiple discovery tasks, including activity prediction, molecular design, target prediction, toxicity assessment, and candidate prioritisation, when selected and interpreted in relation to expert-defined constraints [4]. The aim of this article is therefore to develop a human-in-the-loop computational phytopharmacology framework that clarifies where human expertise is needed across data curation, model selection, output interpretation, mechanistic plausibility assessment, evidence grading, safety review, translational decision confidence, and validation planning. The framework does not claim that expert review proves computational outputs correct; rather, it proposes structured expert involvement as a documented, uncertainty-aware, and validation-oriented safeguard.

Human-in-the-loop phytopharmacology

Human-in-the-loop computational phytopharmacology can be defined as a structured approach in which domain experts contribute before, during, and after computational inference. In molecular design, human-in-the-loop workflows have shown that expert interaction can be embedded within model-guided generation rather than applied only after outputs are produced [5]. Translated to phytopharmacology, this means that experts should not merely inspect final target lists, docking outputs, pathway diagrams, or candidate rankings; they should also shape the curation rules, modelling assumptions, decision boundaries, and validation priorities that determine how those outputs are generated and used.

Human expert feedback may also contribute to iterative model refinement when data are limited or goals are difficult to encode fully in model objectives. Human-in-the-loop active learning for molecule generation illustrates how expert input can augment model-guided search and help refine computational prioritisation during the modelling process [6]. In phytopharmacology, an analogous role may involve expert review of natural-product chemical space, biologically plausible target classes, assay relevance, exposure constraints, safety concerns, or medicinal-chemistry feasibility before further computational cycles are undertaken.

Human-in-the-loop design also requires attention to workflow, accountability, evaluation, and monitoring. Clinical AI scholarship has shown that algorithmic systems require careful integration into human workflows, validation pathways, and governance structures before they can responsibly affect high-stakes decisions [7]. Although computational phytopharmacology usually operates earlier in the research pipeline, the same principle applies: computational outputs should be linked to explicit review points that define whether an output supports exploratory ranking, mechanistic hypothesis generation, validation planning, or no further action.

Explainability is one important component of human-in-the-loop interpretation, but it should not be mistaken for validation. Healthcare explainability scholarship emphasises that explanations are context-dependent and should support human understanding, communication, and scrutiny rather than automatically establish correctness [8]. In computational phytopharmacology, an explanation for a target prediction, pathway association, feature attribution, or toxicity alert may help experts interrogate a model, but it does not prove that a phytochemical, extract, target, or pathway has causal therapeutic relevance.

Expert curation and model oversight

Expert curation begins before model inference. Curated bioactivity resources show the importance of structured assay metadata, target information, activity records, and deposition standards for computational reuse [9]. In human-in-the-loop phytopharmacology, expert curation should therefore assess whether bioactivity evidence is traceable, whether assay endpoints are relevant to the biological claim, whether controls and concentrations are interpretable, and whether labels used for modelling reflect reliable pharmacological evidence rather than noisy or context-mismatched observations.

Chemical annotation is another critical pre-inference checkpoint. Natural product databases provide valuable chemical-structure resources for computational research, but their use in modelling still requires review of compound identity, structure representation, stereochemical ambiguity, duplicate entries, source attribution, and chemical-space coverage [10]. For herbal and traditional medicine contexts, curated herb–ingredient–target–disease resources can support evidence integration, but expert review is still needed to judge whether botanical identity, ingredient evidence, target associations, and disease links are sufficiently coherent for mechanistic hypothesis generation [11].

Model oversight continues during inference through review of annotation pipelines, training data, model selection, applicability domain, output ranking, uncertainty, and potential bias. Molecular networking and metabolome annotation workflows can enhance natural-product interpretation, yet annotation outputs remain dependent on data quality, tool assumptions, and confidence-aware interpretation [12]. Human oversight should therefore examine whether model inputs are appropriate, whether the computational method matches the research question, whether outputs are overextended beyond their evidence base, and whether uncertainty is communicated in a way that supports cautious prioritisation rather than false precision.

Post-inference oversight connects computational outputs to validation planning. Best-practice guidance in phytopharmacological research highlights the need to address methodological quality, assay relevance, reproducibility, and biological interpretation when evaluating natural-product evidence [13]. For extract-based work, transparent chemical characterisation is also essential because pharmacological or toxicological interpretation depends on knowing what material was tested and how it was characterised [14]. **Table 1** summarises expert curation and model oversight functions required for human-in-the-loop computational phytopharmacology.

Table 1. Expert Curation and Model Oversight Functions in Human-in-the-Loop Computational Phytopharmacology: Botanical Identity, Chemical Annotation, Bioactivity Evidence, Assay Relevance, Training Data, Model Outputs, Explainability, Uncertainty, Bias, and Validation Needs

Human-in-the-loop function	Core expert question	Evidence reviewed	Model or data element affected	Uncertainty or bias addressed	Decision-confidence contribution	Validation need	Failure risk if absent
Botanical identity review	Is the plant material taxonomically credible and traceable?	Botanical name, synonym status, source metadata, voucher or authentication information	Herb identity labels, extract metadata, source mapping	Misidentification, synonym confusion, source ambiguity	Prevents false attribution of activity to the wrong botanical source	Independent botanical verification and transparent reporting	Incorrect species identity drives misleading compound, target, or safety interpretation
Chemical annotation review	Are compounds or extract constituents annotated with appropriate confidence?	Spectral evidence, structure records, annotation level, stereochemistry, duplicate structures	Compound lists, molecular features, chemical-space inputs	False annotations, isomer ambiguity, duplicate entries, incomplete structures	Improves reliability of chemical inputs used for prediction and ranking	Orthogonal chemical confirmation where needed	False compound identity leads to unreliable target, docking, ADMET, or pathway inference
Bioactivity evidence review	Does the activity evidence support the modelling label or claim?	Assay endpoint, concentration, controls, target context, activity threshold	Training labels, activity matrices, candidate ranking	Noisy labels, assay heterogeneity, weak activity evidence	Separates stronger pharmacological evidence from weak or ambiguous observations	Replication or independent assay confirmation	Model learns unreliable activity patterns or overweights weak evidence
Assay relevance assessment	Does the assay match the biological or disease-related question?	Cell type, organism, target format, disease relevance, endpoint interpretation	Assay inclusion, endpoint weighting, evidence grade	Context mismatch, endpoint overinterpretation, biological non-equivalence	Defines whether evidence supports mechanistic plausibility or only exploratory screening	Fit-for-purpose pharmacological validation	Irrelevant assay evidence is interpreted as therapeutic relevance

Target evidence review	Are predicted or curated targets supported by credible evidence?	Target annotations, binding evidence, activity evidence, disease linkage	Target prediction outputs, network nodes, pathway inputs	Database incompleteness, target misannotation, hub bias	Strengthens target-level plausibility assessment	Experimental target engagement or perturbation testing	Weak target associations produce misleading mechanism claims
Pathway interpretation review	Are pathway findings biologically coherent?	Enrichment outputs, pathway databases, target distribution, disease biology	Pathway maps, mechanism hypotheses, network interpretation	Overrepresentation bias, hub-driven enrichment, pathway redundancy	Converts pathway outputs into bounded mechanistic hypotheses	Mechanistic validation using appropriate biological systems	Pathway association is mistaken for causal mechanism
Training-data oversight	Are the training data suitable for the intended task?	Dataset provenance, chemical coverage, label quality, assay balance, exclusions	Model training set, feature space, label definitions	Distribution shift, class imbalance, missing negative data	Sets the pre-inference confidence boundary for model use	External testing and dataset audit	Model is applied outside its valid chemical or biological domain
Model selection oversight	Is the selected model appropriate for the question and evidence base?	Model assumptions, endpoint type, interpretability needs, data volume	Algorithm choice, modelling workflow, evaluation plan	Model-task mismatch, unnecessary complexity, black-box overuse	Aligns computation with the research decision being supported	Benchmarking and sensitivity analysis	Sophisticated model is used for an unsuitable endpoint or weak dataset
Model-output interpretation	What decision, if any, can be supported by the output?	Scores, rankings, confidence indicators, domain constraints	Candidate prioritisation, target lists, pathway claims	False precision, score inflation, out-of-domain predictions	Limits output use to research prioritisation or validation planning	Prospective or retrospective validation	High score is treated as proof of activity or mechanism
Explainability assessment	Does the explanation clarify or conflict with domain knowledge?	Feature attribution, structural alerts, pathway explanations, model rationale	Explanation layer, interpretation record, reviewer notes	Post hoc rationalisation, misleading attribution	Supports expert interrogation without implying causal proof	Experimental or mechanistic follow-up	Explanation is treated as validation
Uncertainty interpretation	How uncertain is the prediction and why?	Applicability domain, calibration evidence, uncertainty estimates, missing data	Confidence category, decision boundary, validation priority	Poor calibration, unrecognised uncertainty, out-of-distribution inputs	Prevents overconfident prioritisation from point estimates alone	Calibration assessment and external validation	Uncertain prediction is advanced as high-confidence evidence
Bias detection	Could the model or dataset systematically favour certain compounds, targets, assays, or source traditions?	Dataset composition, source distribution, assay imbalance, publication bias indicators	Dataset weighting, interpretation caveats, feedback priorities	Selection bias, publication bias, chemical-space bias	Adds caution to evidence grading and prioritisation	Bias audit and sensitivity analysis	Framework reproduces distorted evidence patterns
Safety and interaction review	Are toxicity or interaction concerns plausible enough to	Toxicity evidence, ADMET predictions, interaction	Safety flags, exclusion criteria, validation plan	Sparse safety evidence, interaction underreporting,	Prevents premature translational interpretation	In vitro, in vivo, or pharmacological safety testing as appropriate	Candidate is advanced despite plausible safety

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

	affect prioritisation?	alerts, exposure context		exposure uncertainty			or interaction concern
Validation planning	What must be tested before the output can support a stronger claim?	Evidence gaps, assay suitability, mechanism hypothesis, safety concerns	Experimental plan, prioritisation boundary, audit trail	Confirmation bias, incomplete evidence chain	Converts computational output into testable research plan	Stage-appropriate chemical, biological, pharmacological, or safety validation	Computational hypothesis remains unsupported and overinterpreted

Mechanistic reasoning and decision confidence

Mechanistic reasoning in human-in-the-loop computational phytopharmacology begins with careful interpretation of target and pathway evidence. Network pharmacology resources can support target, pathway, and disease-association analysis, but their outputs depend on database completeness, target annotation quality, and assumptions used to connect compounds, herbs, proteins, and disease terms [15]. Expert review is therefore needed to determine whether a predicted target or enriched pathway represents a biologically plausible hypothesis, a database artefact, a hub-driven signal, or an association that is too weak for further prioritisation.

Mechanistic interpretation should also distinguish plausible explanation from causal proof. Network pharmacology can help organise multi-target and systems-level reasoning, but pathway-level associations require cautious interpretation because computational linkage alone does not establish that a compound, extract, or phytochemical mixture produces a therapeutic effect through that pathway [16]. Human experts should therefore evaluate whether the proposed mechanism is coherent across compound identity, target evidence, pathway context, phenotype relevance, exposure plausibility, and available pharmacological evidence.

Molecular docking and related structure-based predictions may contribute to target plausibility assessment, but they require expert interpretation of scoring functions, binding-site assumptions, protein preparation, ligand representation, and biological relevance. Docking can support early-stage ligand–target hypothesis generation, yet docking outputs should be treated as plausibility evidence rather than direct evidence of bioactivity or therapeutic mechanism [17]. Within a human-in-the-loop framework, docking evidence should be reviewed alongside chemical annotation confidence, assay evidence, target biology, exposure feasibility, and validation needs.

Decision confidence in this framework is a bounded research judgement rather than a numeric guarantee of correctness. Probabilistic modelling perspectives in drug discovery show that uncertainty arises from data, model assumptions, biological variability, and incomplete knowledge, so decision confidence should incorporate uncertainty rather than rely on point predictions alone [18]. Reviews of uncertainty quantification in AI-enabled drug discovery further support the need to communicate uncertainty when interpreting model trust and prediction reliability [19]. Explainable AI can help experts interrogate molecular features, predicted mechanisms, or model rationales, but explanation should support expert review rather than be treated as causal validation [20].

Translational review points

Translational review points define where computational outputs are examined before they are used to support stronger research claims. Safety and interaction review is one such checkpoint because graph-based modelling of polypharmacy side effects demonstrates how computational systems can identify possible interaction patterns that require expert safety interpretation before practical use [21]. In phytopharmacology, this checkpoint should include toxicity evidence, ADMET concerns, herb–drug interaction plausibility, exposure uncertainty, and patient-context caution without converting computational alerts into clinical recommendations.

Prediction reliability also requires review of calibration and uncertainty. Calibration research in predictive analytics shows that poorly calibrated predictions can appear confident while failing to align with observed risk, which makes calibration review important for responsible decision-confidence assessment [22]. For computational phytopharmacology, this means that ranked candidates, pathway scores, target probabilities, toxicity alerts, or activity predictions should be interpreted through uncertainty, applicability-domain, and calibration evidence when such evidence is available.

Translational readiness should be separated from discovery-stage prioritisation. Staged evaluation frameworks for AI in medicine emphasise that algorithmic systems require different levels of evidence across development, validation, implementation, and clinical evaluation [23]. By analogy, a phytopharmacology model output may justify research prioritisation or validation planning, but it should not be framed as clinical readiness, therapeutic recommendation, or regulatory acceptability unless supported by appropriate evidence outside the conceptual framework.

Decision-boundary review is especially important when computational outputs could be overinterpreted.

Arguments for interpretable models in high-stakes decisions support the principle that opaque or weakly justified outputs should not be allowed to carry decision weight beyond their evidentiary limits [24]. In human-in-the-loop computational phytopharmacology, the decision boundary should specify whether the evidence supports data correction, model refinement, mechanistic hypothesis generation, experimental validation, translational review, or no action. **Figure 1** illustrates human-in-the-loop translational review points that connect expert curation, model oversight, mechanistic reasoning, safety review, validation planning, and decision-confidence boundaries.

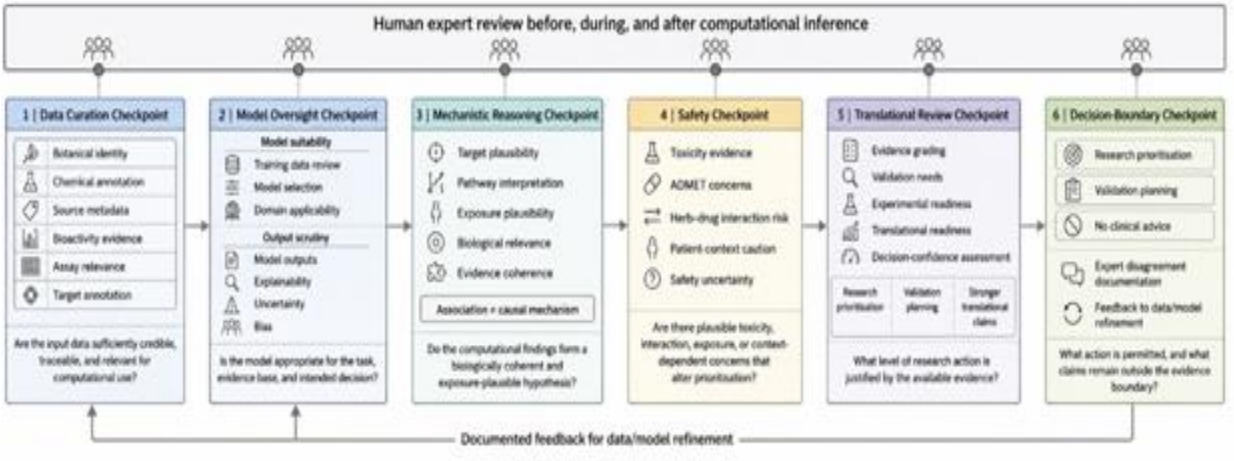

Figure 1. Human-in-the-Loop Translational Review Points in Computational Phytopharmacology

Alt-text description

A checkpoint diagram showing human experts reviewing botanical identity, chemical annotation, bioactivity evidence, assay relevance, model outputs, explainability, uncertainty, mechanistic plausibility, safety concerns, validation needs, translational readiness, and decision boundaries.

Proposed human-in-the-loop framework

The proposed framework defines human-in-the-loop computational phytopharmacology as a structured, documented, and validation-aware research process. Governance scholarship in healthcare AI supports the need for explicit roles, accountability structures, transparency, and decision boundaries when AI systems are used in consequential biomedical contexts [25]. Applied to phytopharmacology, the proposed framework assigns expert responsibilities across botanical identity review, chemical annotation review, bioactivity evidence review, assay relevance assessment, model selection oversight, model-output interpretation, uncertainty review, mechanistic reasoning, safety review, evidence grading, validation planning, and translational decision-boundary control.

The first stage is expert curation, where botanical, chemical, pharmacological, assay, target, pathway, and safety evidence are checked before computational inference. The second stage is computational modelling, where target prediction, pathway inference, ADMET prediction, knowledge-graph analysis, network pharmacology, docking, or candidate prioritisation may be used within explicit task boundaries. Responsible machine-learning roadmaps in healthcare emphasise lifecycle monitoring, bias evaluation, and feedback loops, which support the inclusion of structured model oversight before, during, and after inference [26].

The third stage is mechanistic reasoning, in which experts judge whether computational outputs form a coherent and biologically plausible evidence chain. Ethical analyses of healthcare AI highlight transparency, accountability, and stakeholder responsibility, which support documentation of expert disagreement, uncertainty, assumptions, and decision rationale in the proposed framework [27]. In phytopharmacology, this documentation should clarify whether a mechanism is supported by compound identity, target evidence, pathway coherence, exposure plausibility, safety interpretation, and validation planning, or whether it remains a weak exploratory association.

The fourth stage is translational review, where evidence grading, safety review, interaction concerns, validation needs, decision confidence, and decision boundaries are assessed before any output is used to justify further action. Dual-use analyses of AI-powered drug discovery show that computational systems can generate risks when model outputs are used without adequate safeguards, making misuse, safety, and boundary review relevant to responsible discovery governance [28]. In this framework, translational decision confidence supports research prioritisation and validation planning only; it does not

establish clinical utility, regulatory acceptability, or therapeutic recommendation.

The final stage is feedback-driven improvement. Expert disagreement, annotation errors, failed validation attempts, uncertainty signals, safety concerns, and model-output limitations should be returned to data curation and model refinement through versioned audit trails. This feedback loop makes human expertise a continuous part of computational phytopharmacology rather than an informal final approval step. **Table 2** presents the proposed human-in-the-loop framework for computational phytopharmacology.

Table 2. Proposed Human-in-the-Loop Framework for Computational Phytopharmacology: Expert Curation, Model Oversight, Mechanistic Reasoning, Safety Review, Evidence Grading, Translational Review Points, Decision Confidence, Feedback Loops, and Validation Boundaries

Framework stage	Core purpose	Human expert role	Computational element reviewed	Evidence required	Decision-confidence contribution	Validation boundary	Feedback action	Failure risk
Expert curation	Establish reliable input evidence before inference	Review botanical identity, chemical annotation, source metadata, bioactivity evidence, assay relevance, target evidence, and safety records	Herb labels, compound structures, assay labels, target annotations, training inputs	Traceable source data, chemical evidence, assay metadata, target context, safety evidence	Sets the minimum evidentiary basis for computational use	Does not validate pharmacological effect	Correct annotations, exclude weak data, document uncertainty	Models are trained or interpreted using unreliable inputs
Computational modelling	Generate bounded research hypotheses	Define modelling question, endpoint, constraints, and permissible output use	Target prediction, pathway inference, ADMET prediction, network pharmacology, docking, candidate ranking	Fit between task, data, model, endpoint, and intended decision	Produces candidate hypotheses for review	Does not establish mechanism, efficacy, safety, or clinical utility	Record model assumptions, applicability limits, and output scope	Computational score is mistaken for evidence of activity
Model oversight	Examine model suitability and output reliability	Review training data, model selection, output ranking, explainability, uncertainty, bias, and applicability domain	Training set, features, predictions, explanations, uncertainty indicators	Dataset provenance, evaluation evidence, uncertainty signals, explanation records	Defines whether outputs are usable for prioritisation, validation planning, or rejection	Requires external testing or prospective validation where claims are strengthened	Update model choice, revise data, flag bias or uncertainty	Overconfident or biased model output guides poor prioritisation
Mechanistic reasoning	Assess biological plausibility and evidence coherence	Interpret compound–target–pathway–phenotype relationships	Predicted targets, enriched pathways, molecular networks, docking outputs, mechanism maps	Target confidence, pathway relevance, exposure plausibility, assay consistency, safety signals	Converts disconnected predictions into bounded mechanism hypotheses	Mechanistic claim requires experimental perturbation or pharmacological validation	Refine mechanism hypothesis and document weak links	Correlation or database association is presented as causal proof
Safety review	Identify toxicity and interaction	Assess toxicity evidence, ADMET alerts, herb–drug interaction plausibility,	Toxicity predictions, interaction networks, safety	Toxicological evidence, interaction evidence, exposure	Adds caution to prioritisation and can halt escalation	Requires fit-for-purpose safety testing	Flag exclusions, prioritise safety validation,	Candidate advances despite plausible safety concern

	concerns before translational escalation	exposure context, and safety uncertainty	flags, ADMET outputs	assumptions, safety uncertainty			document uncertainty	
Evidence grading	Integrate curation, modelling, mechanism, uncertainty, and safety evidence	Assign qualitative evidence category and rationale without inventing numeric certainty	Evidence chain, reviewer notes, uncertainty profile, validation plan	Data quality, assay relevance, model reliability, mechanism plausibility, safety review	Makes confidence transparent and bounded	Evidence grade does not equal clinical readiness	Record rationale, disagreement, missing evidence, and next tests	Selective or unbalanced evidence supports overclaiming
Translational review	Determine whether output supports prioritisation, validation, or no action	Review readiness, feasibility, uncertainty, validation requirements, and decision boundary	Candidate recommendation, mechanism hypothesis, safety flags, validation plan	Evidence grade, uncertainty, safety review, feasibility, validation pathway	Separates research prioritisation from stronger translational claims	Clinical or regulatory use requires separate evidence and review	Escalate to validation, revise model, or stop candidate	Discovery hypothesis is framed as therapeutic recommendation
Decision-confidence assessment	Define the permitted decision from available evidence	Judge confidence as bounded, qualitative, and validation-aware	Confidence statement, decision log, evidence grade, uncertainty record	Complete audit trail, reviewer rationale, disagreement record, validation needs	Supports transparent research decisions without implying certainty	Does not guarantee correctness or clinical utility	Update decision boundary after new evidence	Confidence language becomes false assurance
Audit trail and versioning	Preserve accountability and reproducibility	Document data versions, model versions, review decisions, uncertainty, disagreement, and changes	Dataset version, model version, reviewer log, evidence record	Version history, review notes, curation changes, validation outcomes when available	Makes decisions traceable and revisable	Documentation is not validation by itself	Maintain versioned records and update after error detection	Later users cannot reconstruct why a decision was made
Feedback-driven improvement	Improve data, models, and review logic iteratively	Convert errors, disagreements, validation findings, and uncertainty into revisions	Data curation rules, model parameters, evidence thresholds, review checkpoints	Error analysis, validation feedback, disagreement patterns, missing data	Strengthens future research prioritisation processes	Improvement requires re-evaluation after changes	Refine datasets, retrain or replace models, revise checkpoints	Repeated errors persist because feedback is not captured

Figure 2 presents the proposed human-in-the-loop framework for computational phytopharmacology, linking expert curation, model oversight, mechanistic reasoning, translational review, and feedback-driven improvement.

Figure 2. Human-in-the-Loop Framework for Computational Phytopharmacology

Alt-text description

A circular human-in-the-loop framework showing expert curation, computational modeling, model oversight, mechanistic reasoning, evidence grading, safety review, translational review, decision-confidence assessment, validation planning, and feedback loops to improve data and models.

CONCLUSION

This article proposes a human-in-the-loop framework for computational phytopharmacology that integrates expert curation, model oversight, mechanistic reasoning, safety review, translational review points, decision-confidence assessment, validation planning, audit trails, and feedback-driven improvement. The framework argues that human expertise should be embedded as a structured, documented, and validation-aware component before, during, and after computational inference rather than added as an informal final check on model outputs. Its purpose is to support responsible research prioritisation by clarifying the evidentiary limits of computational predictions, the expert questions required at each checkpoint, and the decision boundaries that separate hypothesis generation from clinical or regulatory claims. The framework remains conceptual and should be interpreted as a methodological contribution for organising expert review, uncertainty communication, and validation planning in computational phytopharmacology.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] 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

[2] 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

[3] 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

[4] Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci. 2019;40(8):592-604. doi:10.1016/j.tips.2019.06.004

[5] Sundin I, Voronov A, Xiao H, Papadopoulos K, Bjerrum EJ, Heinonen M, et al. Human-in-the-loop assisted de novo molecular design. J Cheminform. 2022;14(1):86. doi:10.1186/s13321-022-00667-8

[6] Nahal Y, Menke J, Martinelli J, Heinonen M, Kabeshov M, Janet JP, et al. Human-in-the-loop active learning for goal-oriented molecule generation. J Cheminform. 2024;16(1):138. doi:10.1186/s13321-024-00924-y

[7] 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

[8] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. 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

[9] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: Towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1):D930-40. doi:10.1093/nar/gky1075

[10] Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: Collection of Open Natural Products database. J Cheminform. 2021;13(1):2. doi:10.1186/s13321-020-00478-9

[11] Fang SS, Dong L, Liu L, Guo JC, Zhao LH, Zhang JY, et al. HERB: A high-throughput experiment- and reference-guided database of traditional Chinese medicine. Nucleic Acids Res. 2021;49(D1):D1197-206. doi:10.1093/nar/gkaa1063

[12] Ernst M, Kang KB, Caraballo-Rodríguez AM, Nothias LF, Wandy J, Chen C, et al. MolNetEnhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools. Metabolites. 2019;9(7):144. doi:10.3390/metabo9070144

[13] Heinrich M, Appendino G, Efferth T, Fürst R, Izzo AA, Kayser O, et al. Best practice in research—overcoming common challenges in phytopharmacological research. J Ethnopharmacol. 2020;246:112230. doi:10.1016/j.jep.2019.112230

[14] Heinrich M, Jalil B, Abdel-Tawab M, Echeverria J, Kulić Ž, McGaw LJ, et al. Best practice in the chemical characterisation of extracts used in pharmacological and toxicological research—the ConPhyMP-Guidelines. Front Pharmacol. 2022;13:953205. doi:10.3389/fphar.2022.953205

[15] Zhang R, Zhu X, Bai H, Ning K. Network pharmacology databases for traditional Chinese medicine: Review and assessment. Front Pharmacol. 2019;10:123. doi:10.3389/fphar.2019.00123

[16] Nogales C, Mamdouh ZM, List M, Kiel C, Casas AI, Schmidt HHHW. Network pharmacology: Curing causal mechanisms instead of treating symptoms. Trends Pharmacol Sci. 2022;43(2):136-50. doi:10.1016/j.tips.2021.11.004

[17] Pinzi L, Rastelli G. Molecular docking: Shifting paradigms in drug discovery. Int J Mol Sci. 2019;20(18):4331. doi:10.3390/ijms20184331

[18] Lazic SE, Williams DP. Quantifying sources of uncertainty in drug discovery predictions with probabilistic models. Artif Intell Life Sci. 2021;1:100004. doi:10.1016/j.ailsci.2021.100004

[19] Yu J, Wang D, Zheng M. Uncertainty quantification: Can we trust artificial intelligence in drug discovery? iScience. 2022;25(8):104814. doi:10.1016/j.isci.2022.104814

[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] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457-66. doi:10.1093/bioinformatics/bty294

[22] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7

[23] Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: Phases of clinical research. JAMIA Open. 2020;3(3):326-31. doi:10.1093/jamiaopen/ooaa033

[24] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-15. doi:10.1038/s42256-019-0048-x

[25] Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc. 2020;27(3):491-7. doi:10.1093/jamia/ocz192

[26] 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

[27] Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med. 2020;260:113172. doi:10.1016/j.socscimed.2020.113172

[28] Urbina F, Lentzos F, Invernizzi C, Ekins S. Dual use of artificial-intelligence-powered drug discovery. Nat Mach Intell. 2022;4(3):189-91. doi:10.1038/s42256-022-00465-9