Responsible AI for Phytopharmaceutical Innovation: Scientific Validity, Ethical Sourcing, Explainability, Reproducibility, and Clinical Trust

Ahmad Razak[1]*, Norazizah Mohamad[1], Fairuz Ismail[2]

[1]Department of AI for Tropical Plant Metabolomics, Faculty of Pharmacy, University of Malaya, Kuala Lumpur, Malaysia.
[2]Department of Computational Pharmacology for Anti-Diabetic Compounds, Faculty of Pharmacy, Universiti Sains Malaysia, Penang, Malaysia.

ABSTRACT

Artificial intelligence is increasingly used to support phytopharmaceutical innovation through natural product discovery, compound prioritization, mechanism interpretation, safety assessment, and translational decision-making. Yet responsible use of AI in this field requires more than technical model performance, because phytopharmaceutical research involves complex biological materials, variable phytochemical composition, heterogeneous evidence, traditional knowledge contexts, biodiversity concerns, and clinically sensitive interpretations. This article develops an original responsible innovation framework for AI-enabled phytopharmaceutical research. The framework connects scientific validity, botanical and phytochemical integrity, data provenance, ethical sourcing, traditional knowledge sensitivity, biodiversity stewardship, explainability, reproducibility, uncertainty communication, evidence grading, human expert review, clinical trust, translational safeguards, and governance accountability. It argues that AI outputs in phytopharmaceutical innovation should be treated as decision-support evidence rather than proof of efficacy, safety, mechanism, ownership, or clinical readiness. Scientific validity is positioned as the foundation of responsible AI, requiring clear research questions, reliable botanical identity, chemical characterization, data quality, and validation planning. Ethical sourcing is treated as inseparable from scientific trust because provenance, cultural context, benefit-sharing awareness, and ecological stewardship shape the legitimacy of data reuse. Explainability and reproducibility are presented as necessary but insufficient safeguards that must be combined with uncertainty communication and expert review. The framework offers a structured governance logic for cautious, transparent, and evidence-aware AI-enabled phytopharmaceutical innovation.

Key Words: *Responsible AI, Phytopharmaceuticals, Natural products, Scientific validity, Ethical sourcing, Explainability*

eIJPPR 2025; 15(5):34-43

HOW TO CITE THIS ARTICLE: Razak A, Mohamad N, Ismail F. Responsible AI for Phytopharmaceutical Innovation: Scientific Validity, Ethical Sourcing, Explainability, Reproducibility, and Clinical Trust. Int J Pharm Phytopharmacol Res. 2025;15(5):34-43. https://doi.org/10.51847/Z3OtH4ranw

INTRODUCTION

AI-enabled phytopharmaceutical innovation is becoming a significant research direction within natural product discovery because computational systems may support compound prioritization, activity prediction, target hypothesis generation, mechanism exploration, and evidence integration across heterogeneous biological and chemical datasets [1].

In this setting, AI should be understood as a research-support capability rather than an autonomous authority, because phytopharmaceutical claims remain dependent on the validity of biological materials, assay contexts, computational assumptions, and downstream experimental confirmation.

Natural product drug discovery remains scientifically productive but methodologically complex, since biological source identity, phytochemical diversity, extract

Corresponding author: Ahmad Razak
Address: Department of AI for Tropical Plant Metabolomics, Faculty of Pharmacy, University of Malaya, Kuala Lumpur, Malaysia.
E-mail: ✉ ahmad.razak@um.edu.my
Received: 18 May 2025; **Revised:** 29 September 2025; **Accepted:** 29 September 2025

variability, bioavailability, safety, and translational feasibility all affect how evidence should be interpreted [2]. These complexities are especially relevant for phytopharmaceutical innovation because an AI model trained on incomplete or poorly characterized data may amplify uncertainty, obscure source variability, or generate apparently precise but biologically fragile predictions.

Responsible AI scholarship has shown that ethical and governance concerns cannot be reduced to technical accuracy, because transparency, accountability, fairness, privacy, human agency, and societal impact shape whether AI systems are legitimate and trustworthy [3]. For phytopharmaceutical innovation, this broader view is essential because computational reuse of natural product data may involve cultural knowledge, biodiversity resources, benefit-sharing expectations, and clinical trust boundaries that cannot be resolved by algorithmic performance alone.

Health-related machine learning requires harm-prevention logic, validation discipline, and careful deployment safeguards before outputs are used to influence clinical or translational decisions [4]. This article therefore proposes a responsible innovation framework for AI-enabled phytopharmaceutical research, with the aim of connecting scientific validity, ethical sourcing, explainability, reproducibility, clinical trust, translational safeguards, and governance accountability into a coherent framework for cautious and transparent innovation.

Responsible AI in phytopharmaceutical innovation

Responsible AI in phytopharmaceutical innovation can be defined as the alignment of computational methods with scientifically valid evidence, ethically legitimate data and material sourcing, transparent model interpretation, reproducible workflows, uncertainty-aware reporting, human oversight, and cautious translational decision boundaries. Biomedical AI ethics highlights that machine-learning systems in medicine raise issues of accountability, data responsibility, bias, transparency, and appropriate human control, which are directly relevant when AI outputs may shape pharmacological interpretation or therapeutic prioritization [5].

Implementation ethics is especially important because model outputs may appear authoritative even when they are based on incomplete datasets, uncertain labels, unrepresentative populations, poorly characterized extracts, or weak validation designs. Ethical implementation of machine learning in health care requires oversight, professional judgment, and accountability rather than simple reliance on automated predictions [6]. In phytopharmaceutical innovation, this means that predicted bioactivity, target association, toxicity risk, or candidate ranking should remain subordinate to domain expertise and staged validation.

Trustworthy AI also requires usability, transparency, explainability, and attention to the needs of users who must interpret model outputs in real decision contexts [7]. For phytopharmaceutical research, this requires explanations that are meaningful to pharmacologists, phytochemists, ethnopharmacologists, toxicologists, clinicians, and governance reviewers, rather than generic feature-attribution graphics detached from botanical identity, extract composition, experimental context, and uncertainty.

Responsible AI should therefore be governed across the lifecycle of data acquisition, model development, interpretation, reporting, validation planning, and translational use. A governance model for AI in health care supports this lifecycle perspective by emphasising accountability, oversight, and structured governance for AI applications rather than treating responsible use as a one-time technical review [8]. In phytopharmaceutical innovation, lifecycle governance is needed because scientific validity, sourcing legitimacy, reproducibility, and clinical trust can fail at different stages of the research pipeline.

Scientific validity and ethical sourcing

Scientific validity is the first condition of responsible AI-enabled phytopharmaceutical innovation. Best-practice phytopharmacological research requires careful attention to study design, biological material identity, preparation conditions, pharmacological relevance, and transparent reporting, because weak upstream evidence can compromise downstream interpretation [9]. AI systems trained on or applied to such evidence should therefore be evaluated not only for computational performance but also for the validity of the biological assumptions embedded in the data.

Botanical and phytochemical integrity are particularly important because plant species, plant part, extraction method, processing, batch composition, storage, and chemical characterization can all influence pharmacological interpretation. Best-practice guidance for chemical characterization of extracts emphasises that pharmacological and toxicological research requires transparent characterization of the tested material [10]. In AI-enabled research, this means that a dataset entry labelled only by a plant name, extract category, or traditional preparation is not equivalent to a well-characterized phytopharmaceutical input suitable for robust computational inference.

Data provenance and data quality are also central to responsible AI because natural product databases vary in scope, curation depth, chemical representation, biological

annotation, and source traceability. Reviews of natural product databases show that researchers must critically assess where data originate, how they are structured, and what limitations affect reuse [11]. Source-linked databases that connect natural products with activity and species information further illustrate why responsible AI systems should preserve source context rather than detach chemical structures from biological origin and evidence provenance [12].

Ethical sourcing extends scientific validity by asking whether biological materials, associated knowledge, and data were accessed, represented, and reused in legitimate ways. Access and benefit-sharing discussions in natural product research show that sourcing cannot be treated as a purely administrative issue because it affects fairness, biodiversity stewardship, and research legitimacy [13]. Protection of Indigenous and local knowledge also requires recognition that knowledge is culturally situated and should not be treated as freely extractable data for computational mining [14]. In this framework, scientific validity and ethical sourcing are connected because unreliable provenance, extractive knowledge practices, and weak documentation can undermine both evidence quality and trust.

Explainability and reproducibility

Explainability is necessary for responsible AI in phytopharmaceutical innovation because researchers need to understand why a model prioritizes a compound, predicts an activity, flags a safety concern, or proposes a mechanism-related hypothesis. Explainable AI in drug discovery can support model interpretation and hypothesis generation, but explanations should not be treated as causal proof of pharmacological mechanism or clinical effect [15]. In phytopharmaceutical research, this distinction is crucial because complex mixtures, pleiotropic biological effects, and uncertain target annotations can make apparently coherent explanations misleading if they are not experimentally evaluated.

The usefulness of explanation depends on the user, decision context, and risk level. A multidisciplinary perspective on healthcare AI explainability shows that explanation should be designed around the needs of affected stakeholders and the decisions being supported [16]. A comprehensive survey of explainability in trustworthy healthcare AI further shows that terminology, design choices, and evaluation strategies must be made explicit if explanations are to support trust rather than simply decorate model outputs [17]. For phytopharmaceutical innovation, this means that explainability should clarify decision boundaries, uncertainty, evidence source, and biological plausibility.

Reproducibility is equally important because computational claims cannot be responsibly reused if datasets, code, preprocessing steps, model configurations, evaluation settings, and reporting choices are opaque. Reproducibility research in machine learning for health has shown persistent gaps in transparent data access, implementation detail, and evaluation practices [18]. For AI-enabled phytopharmaceutical research, reproducibility should include versioned data where feasible, transparent workflow documentation, model reporting, code availability when possible, and clear limits on reuse when data are restricted by licensing, ethics, or traditional knowledge governance.

Reproducibility also does not guarantee clinical validity. Healthcare machine-learning models may fail to reproduce because of dataset shift, implementation differences, hidden preprocessing choices, and context-specific evaluation conditions [19]. Digital medicine scholarship similarly connects reproducibility deficits with weakened trust in computational claims [20]. Therefore, explainability and reproducibility should be treated as responsible innovation safeguards that make AI outputs inspectable and repeatable, not as substitutes for experimental validation, safety assessment, clinical evaluation, or regulatory judgment.

Table 1 summarises responsible AI requirements for scientific validity, ethical sourcing, explainability, and reproducibility in phytopharmaceutical innovation.

Table 1. Responsible AI Requirements for Phytopharmaceutical Innovation: Scientific Validity, Botanical and Phytochemical Integrity, Ethical Sourcing, Traditional Knowledge Sensitivity, Explainability, Uncertainty Communication, Reproducibility, Evidence Grading, and Governance Accountability

Responsible AI requirement	Core question	Phytopharmaceutical-specific concern	Required evidence or documentation	AI-system implication	Failure risk if absent	Governance or validation need	Research decision relevance
Scientific validity	Is the research question biologically and pharmacologically justified?	AI outputs may be built on weak assay logic, unclear endpoints, or unsupported	Clear hypothesis, pharmacological rationale, assay context, model purpose,	Models should be aligned with defined scientific questions	Speculative candidate prioritization or unsupported pharmacological interpretation.	Scientific review before modelling and before translational interpretation.	Determines whether computational outputs can support

		mechanism assumptions.	and validation plan.	rather than exploratory claims alone.			research prioritization.
Botanical and phytochemical integrity	Is the biological material sufficiently identified and characterized?	Species, plant part, extraction, processing, batch, and chemical profile may alter interpretation.	Botanical identity, plant part, preparation method, extraction context, batch information, and chemical characterization.	Data labels should preserve botanical and compositional specificity.	Misleading inference from poorly defined extracts or inconsistent materials.	Botanical and phytochemical documentation gate.	Determines whether a dataset entry is fit for AI reuse.
Data provenance	Where did the data originate and how were they transformed?	Natural product data may be detached from source species, assay conditions, or curation history.	Source record, curation method, transformation log, inclusion criteria, and data limitations.	AI pipelines should retain traceability from source evidence to model output.	Irreproducible or unaccountable data reuse.	Provenance audit and version control.	Determines whether outputs can be traced and independently appraised.
Ethical sourcing	Were materials and associated data accessed responsibly?	Biological resources may involve access permissions, benefit-sharing expectations, and biodiversity concerns.	Access context, sourcing route, permission status where applicable, and benefit-sharing awareness.	AI systems should not treat all available data as ethically unrestricted.	Extractive innovation, legitimacy loss, and governance failure.	Ethical sourcing review and documentation.	Determines whether research materials and data are legitimate for use.
Traditional knowledge sensitivity	Is knowledge treated as culturally situated and governed?	Traditional use information may reflect community knowledge, cultural context, and restrictions on reuse.	Knowledge provenance, consent awareness, cultural context, and restrictions where known.	AI data ingestion should include safeguards for sensitive knowledge.	Misappropriation, decontextualization, and overclaiming from traditional use.	Stakeholder-sensitive governance and responsible reporting.	Determines whether traditional knowledge can be ethically represented.
Biodiversity stewardship	Does the research account for ecological responsibility?	AI-driven interest in a species may increase collection or supply-chain pressure.	Conservation awareness, sourcing sustainability, supply-chain context, and ecological risk notes.	Candidate ranking should not ignore ecological vulnerability.	Unsustainable sourcing or biodiversity harm.	Biodiversity risk screening.	Determines whether innovation pathways are ecologically responsible.
Explainability	Can model reasoning be inspected in a decision-relevant way?	Feature importance or pathway labels may be mistaken for confirmed mechanisms.	Explanation method, intended user, interpretation limits, and uncertainty statement.	Explanations should support expert review without replacing validation.	False confidence in predicted mechanisms.	Explanation evaluation and expert interpretation.	Determines whether outputs are interpretable enough for research use.
Uncertainty communication	Are confidence limits and unknowns explicit?	Sparse evidence, extract variability, and dataset bias can create hidden uncertainty.	Confidence indicators, sensitivity checks, known limitations, and	AI interfaces and reports should display uncertainty	Premature translation or overconfident claims.	Uncertainty reporting requirement.	Determines how cautiously outputs

			unknown-risk statements.	alongside predictions.			should be interpreted.
Reproducibility	Can the computational workflow be repeated or checked?	Natural product datasets may change, and restricted data may limit full reproducibility.	Workflow documentation, data versioning, preprocessing steps, model configuration, and code availability where feasible.	Models should be documented for independent appraisal or constrained reuse.	Non-replicable findings and weak trust.	Reproducibility checklist and independent verification where feasible.	Determines whether claims can be tested or reused responsibly.
Evidence grading	What level of support exists for each claim?	In silico prediction, in vitro activity, animal evidence, traditional use, and clinical evidence differ in strength.	Evidence level, validation stage, safety context, and claim boundary.	AI systems should distinguish hypothesis generation from validated evidence.	Evidence inflation and clinical overreach.	Evidence hierarchy and claim-control review.	Determines whether outputs support discovery, validation planning, or translation.
Governance accountability	Who is responsible for decisions, documentation, and failures?	Multi-source data, traditional knowledge, and translational claims involve multiple accountability points.	Roles, review responsibilities, audit trail, escalation route, and reporting obligations.	AI systems should be embedded in accountable research governance.	Responsibility gaps and unreviewed deployment.	Lifecycle governance and auditability.	Determines whether the research process is governable and reviewable.

Clinical trust and translational safeguards

Clinical trust in AI-enabled phytopharmaceutical innovation should be understood as a cautious, evidence-based relationship between computational outputs, experimental validation, safety evidence, expert interpretation, patient protection, and decision boundaries. Key challenges for clinical AI show that technical performance alone is insufficient for clinical impact when real-world implementation, safety, workflow integration, and evidence of benefit remain unresolved [21]. In phytopharmaceutical research, this means that a predicted bioactivity profile, target interaction, safety signal, or candidate ranking should not be interpreted as clinical readiness without staged validation.

Translational safeguards should require AI-supported claims to be assessed for transparency, replicability, ethics, and effectiveness before they are used to justify higher-risk decisions. Patient-benefit AI research has been framed around critical questions concerning transparency, replicability, ethical design, and effectiveness, which are directly relevant to evidence grading in computational pharmacology [22]. For phytopharmaceutical innovation, these questions should be adapted to distinguish discovery support, mechanistic hypothesis generation, preclinical prioritization, safety screening, and clinical decision support.

Responsible reporting is another translational safeguard because incomplete descriptions of AI interventions can prevent meaningful appraisal, replication, and clinical interpretation. AI-specific clinical trial reporting guidance emphasises the need to describe AI interventions transparently when they are evaluated in trial contexts [23]. Although the proposed framework is not a clinical trial guideline, it adopts this reporting logic by requiring clear separation between computational prediction, experimental validation, safety evidence, clinical evidence, and decision-support boundaries.

Early-stage evaluation is especially important for AI-enabled decision-support systems because premature deployment may create clinical, ethical, and governance risks. Reporting guidance for early clinical evaluation of AI decision support highlights the need for staged assessment before wider use [24]. In phytopharmaceutical innovation, translational safeguards should therefore include evidence grading, uncertainty communication, expert review, validation gates, safety awareness, pharmacovigilance sensitivity, and explicit limits on clinical interpretation.

Proposed responsible innovation framework

The proposed responsible innovation framework translates AI ethics and biomedical governance principles into domain-specific requirements for AI-enabled phytopharmaceutical innovation. Reviews of practical AI ethics tools show that responsible AI requires operational methods that connect principles with implementation

practices, documentation, review, and accountability [25]. In this framework, responsible innovation begins before model development, because botanical identity, phytochemical characterization, provenance, sourcing legitimacy, and evidence quality determine whether computational outputs can be interpreted responsibly.

The framework rejects the assumption that ethical principles alone are sufficient to govern AI-enabled phytopharmaceutical research. Critiques of principle-based AI ethics show that high-level principles require institutional structures, practical translation, and accountability mechanisms before they can shape responsible practice [26]. Accordingly, the proposed framework treats scientific validity, ethical sourcing, traditional knowledge sensitivity, biodiversity stewardship, explainability, reproducibility, clinical trust, and governance accountability as operational requirements rather than abstract values.

A central component of the framework is provenance-aware data stewardship. Dataset documentation work has shown that structured records of dataset motivation, composition, collection process, recommended use, and limitations can improve transparency and accountability [27]. Adapted to phytopharmaceutical innovation, this logic requires documentation of botanical source, phytochemical characterization, evidence origin, traditional knowledge context where relevant, data transformation, access constraints, uncertainty, and intended decision boundary.

Explainability is positioned as a bounded governance function rather than a guarantee of mechanistic truth. Explainable AI taxonomies show that interpretability methods can support transparency and responsible use but must be matched to context, users, and limitations. Therefore, the framework requires explanation outputs to be accompanied by uncertainty communication, evidence grading, expert interpretation, and warnings against treating model explanations as confirmed pharmacological mechanisms.

Trustworthy AI combines technical reliability, explainability, accountability, and human acceptance, but these elements must be embedded in lifecycle governance to support responsible use. Recent healthcare AI governance work further supports the need for accountable structures that guide AI adoption, monitoring, oversight, and risk management across implementation contexts. The proposed framework therefore connects data acquisition, model development, evidence interpretation, validation planning, reporting, and translational decision-making through governance checkpoints rather than presenting AI as an independent source of clinical authority.

Table 2 presents the proposed responsible innovation framework for AI-enabled phytopharmaceutical research and translation.

Table 2. Proposed Responsible Innovation Framework for AI-Enabled Phytopharmaceutical Innovation: Scientific Validity, Ethical Sourcing, Explainability, Reproducibility, Clinical Trust, Translational Safeguards, Human Oversight, and Governance Accountability

Framework component	Core purpose	Required practice	Evidence or documentation needed	Trust-building function	Translational safeguard	Failure risk	Implementation priority
Scientific validity gate	Ensure that AI work begins from a justified pharmacological question.	Define research question, endpoint, biological rationale, model purpose, and validation pathway.	Study rationale, assay context, dataset inclusion logic, and validation plan.	Prevents speculative modelling from being framed as evidence.	Blocks premature movement from prediction to claim.	Unsupported candidate prioritization or weak mechanism claims.	Essential at project initiation.
Botanical and phytochemical integrity gate	Ensure that biological inputs are identifiable and chemically interpretable.	Document species, plant part, preparation, extraction, batch, and chemical characterization.	Botanical record, preparation record, phytochemical profile, and material description.	Makes computational inputs interpretable to phytochemistry and pharmacology experts.	Prevents translation of results from poorly defined materials.	Misleading inference from ambiguous extracts or inconsistent compositions.	Essential before dataset construction.
Data provenance and documentation layer	Preserve traceability from source evidence to model output.	Record data source, curation, transformation, versioning,	Provenance record, dataset documentation, preprocessing log, and	Enables auditability and independent appraisal.	Prevents opaque reuse of heterogeneous evidence.	Irreproducible claims and unaccountable data reuse.	Essential throughout data lifecycle.

		restrictions, and intended use.	limitation statement.				
Ethical sourcing and knowledge-sensitivity gate	Protect legitimacy of materials, knowledge, and data reuse.	Assess access context, benefit-sharing awareness, traditional knowledge sensitivity, and biodiversity stewardship.	Sourcing notes, access context, knowledge provenance, restrictions, and stewardship considerations.	Builds ethical legitimacy alongside scientific credibility.	Prevents extractive or culturally insensitive innovation pathways.	Misappropriation, governance failure, and loss of community trust.	Essential before data ingestion and publication.
Model validation and uncertainty layer	Align computational performance with evidence limits.	Use appropriate validation, sensitivity analysis, uncertainty reporting, and decision-boundary statements.	Validation design, performance context, uncertainty statement, and failure-mode analysis.	Prevents overconfidence in model outputs.	Restricts outputs to research-support use unless further evidence exists.	Inflated claims from internally valid but externally weak models.	High priority during model development.
Explainability and expert-interpretation layer	Make model outputs inspectable and reviewable.	Provide interpretable outputs, explanation limits, user-specific interpretation, and expert review.	Explanation method, intended user, interpretation guide, and expert-review record.	Helps researchers understand why outputs were generated.	Prevents explanations from being treated as causal proof.	False mechanistic confidence and automation bias.	High priority during interpretation.
Reproducibility and open-methods layer	Make computational work repeatable or independently appraisable.	Report methods, parameters, code where feasible, data versions, workflow steps, and constraints.	Workflow documentation, code availability statement, version records, and reproducibility checklist.	Supports verification and responsible reuse.	Prevents translation from non-replicable computational findings.	Non-reproducible outputs and weak scientific trust.	High priority before dissemination.
Evidence grading layer	Match claim strength to evidence type.	Distinguish in silico prediction, in vitro evidence, in vivo evidence, safety data, traditional use, and clinical evidence.	Evidence table, claim boundary, validation status, and safety context.	Clarifies what the evidence can and cannot support.	Prevents clinical overreach from early-stage evidence.	Evidence inflation and inappropriate therapeutic inference.	Essential before manuscript or translational reporting.
Human oversight and stakeholder participation layer	Keep AI outputs subordinate to expert and stakeholder review.	Include pharmacological, phytochemical, clinical, ethical, and stakeholder-sensitive review.	Review roles, decision record, stakeholder input where applicable, and escalation pathway.	Reduces automation bias and improves legitimacy.	Ensures risky outputs are reviewed before action.	Unreviewed decisions and social or clinical harm.	High priority for higher-risk projects.
Governance accountability and lifecycle monitoring layer	Assign responsibility across the full AI research lifecycle.	Establish accountability, auditability, responsible reporting, risk	Governance plan, audit trail, risk register, reporting checklist, and	Creates durable responsibility beyond model publication.	Prevents ungoverned reuse or deployment drift.	Responsibility gaps, uncontrolled reuse, and trust erosion.	Essential for translational or decision-support use.

	monitoring, and update procedures.	monitoring plan.

Figure 1 illustrates the proposed responsible innovation framework linking scientific validity, ethical sourcing, explainability, reproducibility, clinical trust, translational safeguards, and governance accountability in AI-enabled phytopharmaceutical innovation.

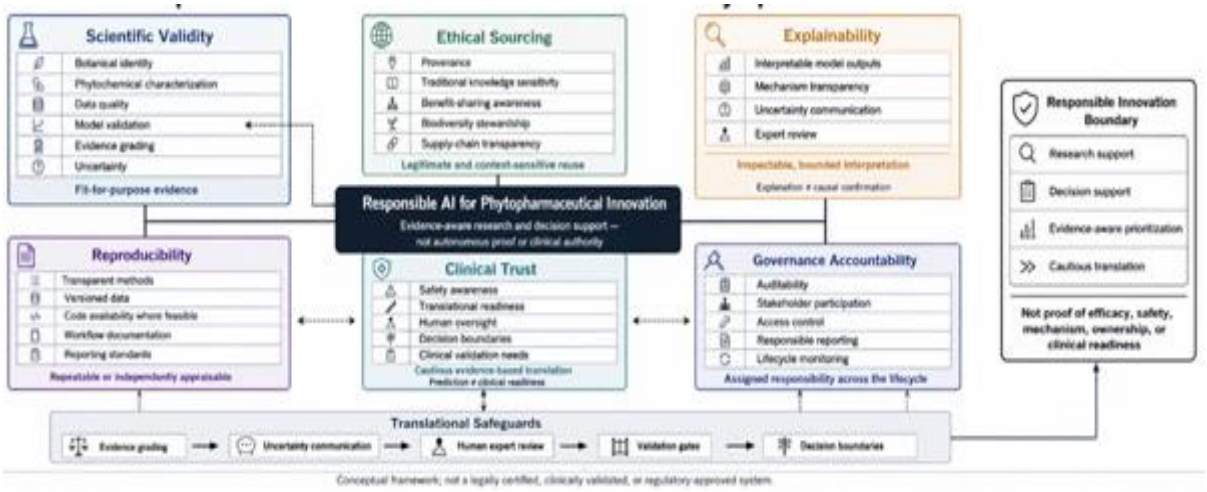

Figure 1. Responsible Innovation Framework for AI-Enabled Phytopharmaceutical Innovation

Alt-text description
A conceptual framework diagram showing responsible AI for phytopharmaceutical innovation. The framework connects scientific validity, ethical sourcing, traditional knowledge sensitivity, biodiversity stewardship, explainability, reproducibility, evidence grading, uncertainty communication, clinical trust, translational safeguards, human expert review, and governance accountability.

CONCLUSION

Responsible AI for phytopharmaceutical innovation requires more than technical model performance. It requires valid botanical and phytochemical inputs, traceable data provenance, ethical sourcing, traditional knowledge sensitivity, biodiversity stewardship, explainable and uncertainty-aware outputs, reproducible workflows, evidence-graded interpretation, human expert review, clinical trust safeguards, translational caution, and accountable governance. The framework proposed in this article positions AI as a research-support and decision-support capability rather than a substitute for pharmacological validation, clinical judgment, ethical review, or regulatory assessment. By connecting scientific validity with ethical legitimacy and translational caution, the framework provides a structured basis for developing AI-enabled phytopharmaceutical research that is transparent, reproducible, context-sensitive, and appropriately bounded.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Sorokina M, Steinbeck C. Review on natural products databases: Where to find data in 2020. J Cheminform. 2020;12(1):20. doi:10.1186/s13321-020-00424-9
[2] Zeng X, Zhang P, He W, Qin C, Chen S, Tao L, et al. NPASS: Natural product activity and species source database for natural product research, discovery and tool development. Nucleic Acids Res. 2018;46(D1):D1217-D22. doi:10.1093/nar/gkx1026
[3] 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
[4] Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E, Gaudry A, et al. The LOTUS initiative for open knowledge management in natural products

research. Elife. 2022;11:e70780. doi:10.7554/eLife.70780

[5] Akpoviri FI, Baharum SN, Zainol ZA. Digital sequence information and the access and benefit-sharing obligation of the Convention on Biological Diversity. Nanoethics. 2023;17(1):1. doi:10.1007/s11569-023-00436-3

[6] Halewood M, Bagley MA, Wyss M, Scholz AH. New benefit-sharing principles for digital sequence information. Science. 2023;382(6670):520-2. doi:10.1126/science.adj1331

[7] Wynberg R, Andersen R, Laird S, Kusena K, Prip C, Westengen OT. Farmers' rights and digital sequence information: Crisis or opportunity to reclaim stewardship over agrobiodiversity? Front Plant Sci. 2021;12:686728. doi:10.3389/fpls.2021.686728

[8] Qin G, Yu H, Wu C. Global governance for digital sequence information on genetic resources: Demand, progress and reforming paths. Glob Policy. 2023;14(2):403-15. doi:10.1111/1758-5899.13202

[9] Sett S, Kress WJ, Halewood M, Nicholson D, Nuñez-Vega G, Faggionato D, et al. Harmonize rules for digital sequence information benefit-sharing across UN frameworks. Nat Commun. 2024;15(1):8745. doi:10.1038/s41467-024-52994-z

[10] Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, et al. The CARE principles for Indigenous data governance. Data Sci J. 2020;19:43. doi:10.5334/dsj-2020-043

[11] Carroll SR, Herczog E, Hudson M, Russell K, Stall S. Operationalizing the CARE and FAIR principles for Indigenous data futures. Sci Data. 2021;8(1):108. doi:10.1038/s41597-021-00892-0

[12] Hudson M, Garrison NA, Sterling R, Caron NR, Fox K, Yracheta J, et al. Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. Nat Rev Genet. 2020;21(6):377-84. doi:10.1038/s41576-020-0228-x

[13] Walter M, Suina M. Indigenous data, Indigenous methodologies and Indigenous data sovereignty. Int J Soc Res Methodol. 2019;22(3):233-43. doi:10.1080/13645579.2018.1531228

[14] Mc Cartney AM, Anderson J, Liggins L, Hudson ML, Anderson MZ, TeAika B, et al. Balancing openness with Indigenous data sovereignty: An opportunity to leave no one behind in the journey to sequence all of life. Proc Natl Acad Sci U S A. 2022;119(4):e2115860119. doi:10.1073/pnas.2115860119

[15] Reyes-García V, Tofighi-Niaki A, Austin BJ, Benyei P, Danielsen F, Fernández-Llamazares Á, et al. Data sovereignty in community-based environmental monitoring: Toward equitable environmental data governance. Bioscience. 2022;72(8):714-7. doi:10.1093/biosci/biac048

[16] Jacobsen A, Azevedo RDM, Juty N, Batista D, Coles S, Cornet R, et al. FAIR principles: Interpretations and implementation considerations. Data Intell. 2020;2(1-2):10-29. doi:10.1162/dint_r_00024

[17] Wilkinson MD, Dumontier M, Sansone SA, Bonino da Silva Santos LO, Prieto M, Batista D, et al. Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci Data. 2019;6(1):174. doi:10.1038/s41597-019-0184-5

[18] Herschel M, Diestelkämper R, Ben Lahmar H. A survey on provenance: What for? What form? What from? VLDB J. 2017;26(6):881-906. doi:10.1007/s00778-017-0486-1

[19] Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé H III, et al. Datasheets for datasets. Commun ACM. 2021;64(12):86-92. doi:10.1145/3458723

[20] McKeown A, Mourby M, Harrison P, Walker S, Sheehan M, Singh I. Ethical issues in consent for the reuse of data in health data platforms. Sci Eng Ethics. 2021;27(1):9. doi:10.1007/s11948-021-00282-0

[21] Tamuhla T, Lulamba ET, Mutemaringa T, Tiffin N. Multiple modes of data sharing can facilitate secondary use of sensitive health data for research. BMJ Glob Health. 2023;8(10):e013092. doi:10.1136/bmjgh-2023-013092

[22] Longpre S, Mahari R, Chen A, Obeng-Marnu N, Sileo D, Brannon W, et al. A large-scale audit of dataset licensing and attribution in AI. Nat Mach Intell. 2024;6(8):975-87. doi:10.1038/s42256-024-00878-8

[23] Pushkarna M, Zaldivar A, Kjartansson O. Data cards: Purposeful and transparent dataset documentation for responsible AI. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency; 2022 Jun 21-24; Seoul, Republic of Korea. New York (NY): Association for Computing Machinery; 2022. p. 1776-826. doi:10.1145/3531146.3533231

[24] Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency; 2021 Mar 3-10; online. New York (NY): Association for Computing Machinery; 2021. p. 610-23. doi:10.1145/3442188.3445922

[25] van der Sloot B, Keymolen E. Can we trust trust-based data governance models? Data Policy. 2022;4:e45. doi:10.1017/dap.2022.36

[26] Paprica PA, Sutherland E, Smith A, Brudno M, Cartagena RG, Crichlow M, et al. Essential requirements for establishing and operating data

trusts: Practical guidance based on a working meeting of fifteen Canadian organizations and initiatives. Int J Popul Data Sci. 2020;5(1):1353. doi:10.23889/ijpds.v5i1.1353

[27] Organisation for Economic Co-operation and Development. Data stewardship, access, sharing and control: A Going Digital III module synthesis report. Paris: OECD Publishing; 2023. Report No.: DSTI/CDEP(2022)6/FINAL.