Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2026, 15(1):75-84
https://doi.org/10.51847/P5SvbmxCBK

Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

FormulationGPT Is Not a Formulator: Governing Generative Artificial Intelligence across Excipient Selection, Process Design, and Product Decisions

David Thompson[1]*, Sarah Mitchell[1], Rachel Adams[2], James Brown[3], Michael Lee[4]

[1]Department of Generative AI Governance in Formulation, Faculty of Pharmacy, University of Toronto, Toronto, Canada.
[2]Department of Excipient Selection and Process Design, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, Canada.
[3]Department of Product Decisions and AI Oversight, Faculty of Pharmacy, McGill University, Montreal, Canada.
[4]Department of FormulationGPT and Clinical Use, Faculty of Pharmacy, University of Guelph, Guelph, Canada.

***Email:** david.thompson@utoronto.ca*

ABSTRACT

Generative artificial intelligence can retrieve, reorganize, and synthesize formulation knowledge, but fluent output is not equivalent to the situated expertise required to design a pharmaceutical product. Formulation decisions depend on dosage-form purpose, material attributes, excipient functionality, compatibility, stability, process conditions, scale, bioavailability, patient needs, and the quality of supporting evidence. This article develops an original, non-empirical governance architecture for generative artificial intelligence used across excipient selection, process design, experimental planning, and product decisions. The architecture separates bounded generative assistance from actions requiring pharmaceutical evidence and accountable authorization. It integrates an intended-use gate, data-provenance and knowledge-grounding controls, formulation and compatibility constraints, process-design and scale-up reasoning, evidence-warrant classification, human review, experiment authorization, product-decision boundaries, and lifecycle monitoring. Uncertainty is treated as a decision-relevant property that must be traced to evidence gaps, model limitations, conflicting sources, or unresolved experimental questions rather than represented by an unqualified confidence statement. Human involvement is defined as an accountable review function with authority to reject, revise, escalate, or authorize proposed work, rather than nominal confirmation of model output. Monitoring and change control extend governance to changes in data, retrieval sources, prompts, models, materials, equipment, and development context. The original contribution is an evidence-to-authorization architecture designed to prevent generated recommendations from crossing directly into pharmaceutical consequences. Its propositions require computational, experimental, manufacturing, human-factors, and implementation validation. The framework does not establish model validity, formulation suitability, clinical benefit, regulatory acceptability, universal applicability, or deployment readiness.

Key words: *Formulation, Pharmaceutical formulation, Dosage form, Excipient, Manufacturability, Stability*
Received: 17 November 2025; **Revised:** 20 February 2026; **Accepted:** 26 February 2026

INTRODUCTION

Pharmaceutical formulation is not the production of plausible combinations of active ingredients, excipients, and process parameters. It is a constrained scientific activity in which composition, dosage form, manufacturing process, stability, release behaviour, bioavailability, patient use, and product-quality requirements must be reconciled. Machine-learning methods can support defined formulation-development tasks, but their outputs remain dependent on the selected representation, training data, development context, and experimental evidence [1].

The growing use of computational models across formulation development has encouraged a broader vision of digital formulation design. Predictive systems may organize prior knowledge, identify patterns, rank candidates, or reduce an initial experimental search space. However, applications differ substantially in task definition, data coverage, interpretability, external validation, and proximity to a consequential product decision [2]. Treating this heterogeneous field as a single capability obscures the difference between assisting scientific inquiry and exercising formulation authority.

Existing formulation platforms illustrate both the promise and the boundary of integrated artificial intelligence. A system may combine literature-derived information, formulation databases, molecular descriptors, and predictive modules to produce apparently coherent recommendations. Such integration can improve access to information and support structured exploration, but it does not demonstrate that the system possesses the contextual judgment, tacit manufacturing knowledge, evidentiary accountability, or decision responsibility of a formulator [3].

This Original Artificial-Intelligence Governance Architecture Article addresses that unresolved boundary. It proposes an evidence-to-authorization architecture for governing generative artificial intelligence across excipient selection, compatibility reasoning, process design, experimental planning, scale-up questions, and product decisions. The contribution is an original conceptual synthesis rather than an empirical validation study, regulatory standard, software qualification report, clinical guideline, or deployment manual. Its purpose is to define governable relationships, testable propositions, validation requirements, failure modes, and explicit limits on decision use.

Why generative output is not equivalent to formulation expertise

Generative output is produced through probabilistic inference over learned or retrieved representations. It may be grammatically fluent, internally persuasive, and technically detailed while containing unsupported statements, source-conflicting claims, omitted constraints, or fabricated connections. Research on natural-language generation identifies hallucination as a multidimensional problem involving factual inconsistency, source disconnection, and unverifiable content [4]. In formulation development, these failures may be especially difficult to detect when incorrect recommendations use credible pharmaceutical terminology.

Domain-relevant knowledge also does not establish comprehensive expert competence. Large language models can encode substantial technical or clinical information and perform strongly on selected benchmarks, yet performance remains conditional on task formulation, evaluation design, prompt context, and the completeness of required knowledge [5]. A formulation decision may require recognition that an apparently reasonable excipient is unsuitable because of concentration, grade, route, residual moisture, processing history, packaging, patient population, or an interaction not represented in the prompt.

Formulation expertise therefore includes more than recall or linguistic synthesis. It combines mechanistic understanding, product-specific evidence, experience with analytical ambiguity, knowledge of manufacturing constraints, interpretation of failed experiments, and responsibility for consequential choices. Ethical analyses of generative artificial intelligence in high-stakes settings similarly indicate that transparency, safety, accountability, and human control cannot be reduced to output accuracy [6]. These principles are relevant because formulation outputs may eventually influence material selection, laboratory work, scale-up, quality strategy, or patient-facing product characteristics.

The distinction is not an assertion that generative systems are intrinsically incapable of useful formulation work. They may expose overlooked alternatives, reorganize dispersed knowledge, identify missing evidence, or assist the drafting of experimental hypotheses. The proposed interpretation is narrower: an output does not become formulation expertise merely because it is detailed, correct in part, or accepted by a user. Expertise is expressed through context-sensitive evaluation and accountable action, whereas generated content remains a proposition whose evidentiary status must be established.

Intended uses and prohibited uses

Governance begins by defining what a system is expected to do. Comparisons between artificial intelligence and professionals can overstate capability when evaluation datasets, reporting practices, user interactions, and real-world consequences do not reproduce professional work [7]. FormulationGPT should therefore not be evaluated as a generalized "formulator." Each intended use should specify the user, dosage-form context, development stage, required inputs, anticipated output, consequence of error, and action that may follow.

Permitted uses may include literature retrieval, source-linked summarization, structured comparison of formulation options, identification of missing information, generation of hypotheses, preparation of draft experimental questions, and organization of known constraints. Conditional uses may include preliminary excipient ranking, interaction-risk flagging, stability-risk triage, formulation-space exploration, or process-scenario generation. These activities require staged evaluation within the intended workflow, including analysis of user interaction, errors, and escalation procedures [8].

Prohibited uses are those in which generated output substitutes for evidence or accountable authorization. These include autonomous declarations of drug–excipient compatibility, final excipient acceptance, release or stability conclusions, bioavailability claims, patient-specific dosage-form recommendations, authorization of process parameters, approval of scale-up, assignment of specifications, batch disposition, and final product decisions. Machine-learning systems can produce unintended consequences that are not visible in conventional performance measures, particularly when users over-trust, misinterpret, or act on outputs outside their evaluated context [9].

This article proposes decision distance as the organizing construct for these use classes. Decision distance describes the unresolved evidentiary and authorization transitions between a generated output and a consequential pharmaceutical commitment. A source-linked summary is relatively distant from product action; an instruction capable of changing an experiment, manufacturing process, specification, or patient-facing product is closer. As decision distance decreases, the required grounding, constraint verification, uncertainty analysis, human authority, and experimental evidence should increase. This classification remains a proposed governance taxonomy and requires empirical evaluation across product types and development environments.

Proposed formulation-AI governance architecture

Responsible artificial intelligence cannot be secured through one model-level control. Governance must extend across problem definition, data construction, development, validation, use, monitoring, and response to failure [10]. The proposed architecture therefore treats formulation-AI governance as a sequence of linked gates through which generated claims may be inspected, constrained, tested, rejected, or authorized. No gate independently establishes product suitability.

The first gate defines the intended user, task, development stage, dosage-form context, allowed output, prohibited action, and consequence of error. Subsequent assessment separates transparency, reproducibility, ethical acceptability, technical performance, operational effectiveness, and scientific benefit because these are distinct questions rather than interchangeable indicators of system quality [11]. A technically reproducible answer may still be unsupported, irrelevant to the product, experimentally false, or unsuitable for decision use.

Responsibility is distributed across the system lifecycle. Data stewards are responsible for source identity and limitations; model developers for system behaviour and version control; formulation scientists for scientific interpretation; analytical, stability, and process specialists for domain-specific evidence; and designated authorities for experiment or product decisions. Ethical analyses of health-related artificial intelligence show that governance requires allocation of responsibility across stakeholders rather than exclusive reliance on technical accuracy [12]. The exact allocation proposed here is formulation-specific conceptual synthesis and must be adapted to organizational roles without implying a predetermined legal or regulatory structure.

The architecture comprises an intended-use and decision-distance gate; model and version identification; data-provenance and knowledge-grounding controls; formulation and compatibility constraints; process-design and scale-up reasoning; uncertainty and evidence-warrant classification; human formulation review; experiment authorization; product-decision boundaries; and lifecycle monitoring. **Figure 1** presents the original conceptual synthesis for governance architecture for generative AI in formulation development, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

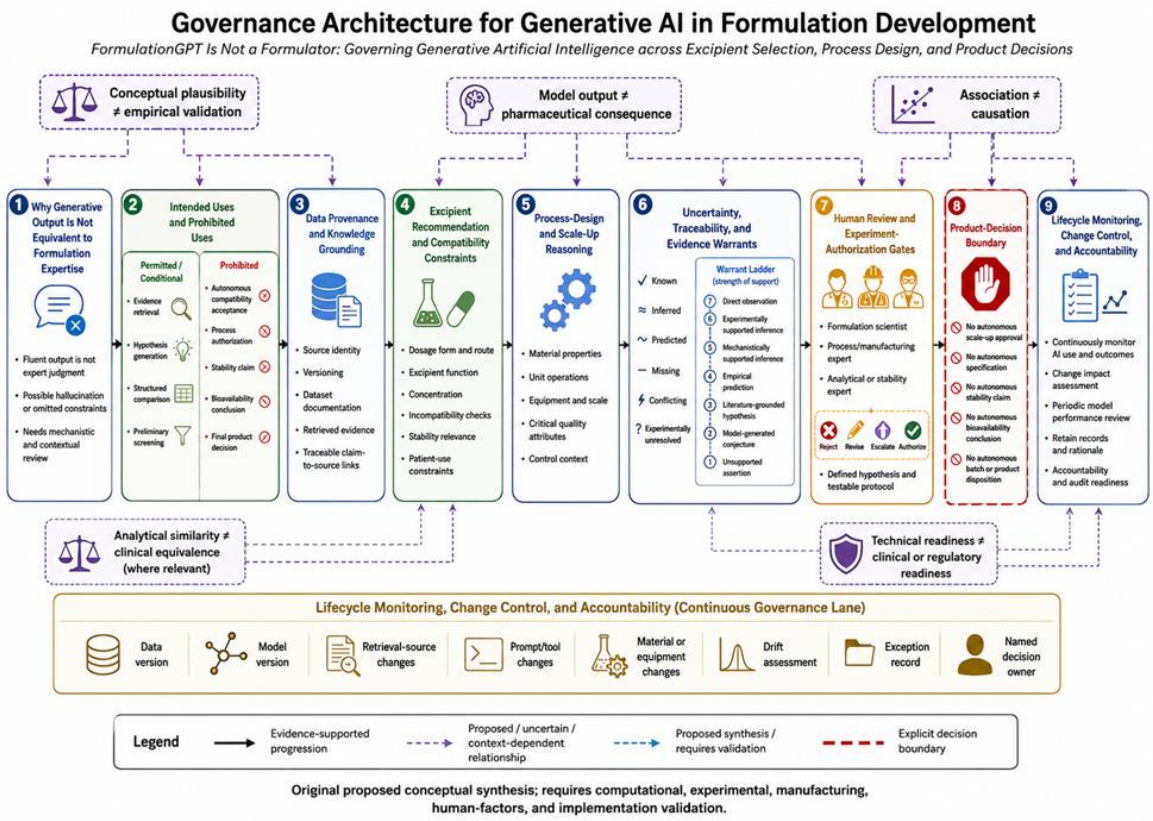

Figure 1. Governance Architecture for Generative AI in Formulation Development.

The figure is an original conceptual synthesis developed for "FormulationGPT Is Not a Formulator: Governing Generative Artificial Intelligence across Excipient Selection, Process Design, and Product Decisions." Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, clinical recommendation, regulatory determination, or deployment-ready system.

The architecture prevents a generative recommendation from moving directly into experimentation or product commitment. Progression requires evidence appropriate to the claim, explicit identification of unresolved uncertainty, verification of formulation and process constraints, and review by a person with relevant competence and authority. Experiment authorization and product authorization are separate boundaries: permission to investigate a generated hypothesis does not validate the hypothesis, and successful experimentation does not automatically establish manufacturing, clinical, or regulatory readiness. The architecture is consequently a proposed organizational design whose scientific validity, usability, failure-detection capacity, and implementation effects remain to be tested.

Data provenance and knowledge grounding

Data provenance should identify where formulation information originated, how it was collected or transformed, its intended use, and its known limitations. Dataset documentation provides a basis for examining composition, context, maintenance, and potential misuse, but documentation alone does not establish scientific validity [13].

Knowledge grounding requires more than attaching a citation to generated text. Structured representations such as knowledge graphs can connect substances, excipient grades, dosage forms, processes, analytical observations, and evidence sources while retaining contextual and provenance relationships [14]. The represented relationships nevertheless remain dependent on source quality and modelling choices.

Retrieval-augmented generation may improve access to external literature and reduce reliance on internal model memory. However, retrieval does not guarantee that a source is correct, complete, applicable to the formulation, or accurately interpreted [15]. Retrieved passages therefore require source verification and product-context assessment.

The architecture proposes a formulation claim ledger recording source identity, version, applicable material and product context, transformation history, evidence type, uncertainty, and reviewer disposition. A claim may be

inspectable without being true, and grounded without being sufficient for compatibility, stability, bioavailability, manufacturability, or product decisions.

Excipient recommendation and compatibility constraints

Drug–excipient compatibility prediction can support early screening when the target outcome, training domain, descriptors, and validation conditions are explicit [16]. A model-generated ranking should therefore be interpreted as candidate prioritization rather than a determination that an excipient is suitable.

Case-validated interaction tools demonstrate that computational screening can be compared with experimental observations, but evidence obtained for selected drug–excipient pairs cannot establish universal compatibility authority [17]. Applicability depends on material grade, concentration, environment, analytical method, processing history, and the interaction definition used.

Physical stability introduces an additional boundary. Machine-learning approaches may predict stability within defined formulation classes, but performance remains conditional on dosage-form representation, data coverage, and experimental confirmation [18]. Chemical compatibility, physical stability, manufacturability, and product performance must not be collapsed into one recommendation score.

The proposed constraint layer applies route, dosage-form, excipient-function, concentration, incompatibility, stability, process, patient-use, supply, and quality constraints before displaying a recommendation. **Table 1** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, components, relationships, and intended uses in formulationgpt is not a formulator.

Table 1. Definitions, components, relationships, and intended uses in FormulationGPT Is Not a Formulator.

Component or scientific dimension	Problem addressed	Inputs or determinants	Proposed mechanism or relationship	Expected contribution	Evidence required	Failure or uncertainty risk	Boundary statement
Generative output	Plausible text may be unsupported	Prompt, model, retrieved content	Output is treated as a reviewable proposition	Faster synthesis and question generation	Source and claim verification	Hallucination, omission, false coherence	Output is not expertise or evidence
Data provenance	Sources may be untraceable	Dataset origin, version, context	Claims remain linked to identifiable evidence [13]	Auditability	Source documentation and applicability review	Hidden transformation or obsolete evidence	Provenance does not prove validity
Knowledge grounding	Evidence may be fragmented	Literature, databases, structured relations	Contextual relations support traceable retrieval [14, 15]	More inspectable reasoning	Retrieval and interpretation testing	Irrelevant or incomplete grounding	Retrieval is not verification
Excipient recommendation	Ranking may be mistaken for suitability	Route, function, grade, concentration, compatibility	Proposed constraints precede candidate ranking [16, 17]	Bounded screening support	Prospective compatibility testing	Unrepresented interaction	Recommendation is not acceptance
Stability reasoning	Predicted stability may not transfer	Composition, process, environment, time	Predictions are linked to formulation class and evidence [18]	Risk prioritization	Product-specific stability studies	Domain shift and missing mechanisms	Prediction is not a stability claim
Process reasoning	Text may ignore equipment and scale	Material, unit operation, equipment, state	Missing process context blocks authorization	Structured scale-up questions	Process and manufacturing evidence	Inapplicable parameters	No autonomous process action

Decision distance	Assistance may move too close to commitment	Intended use and consequence of error	**Proposed:** controls intensify near consequential action	Proportionate governance	Prospective workflow evaluation	Under-classified use	No direct model-to-product transition
Human authorization	Nominal review may lack authority	Expertise, evidence access, responsibility	**Proposed:** reviewers may reject, revise, escalate, or authorize	Accountable action	Human-factors and implementation studies	Automation bias or rubber-stamping	Human presence alone is insufficient

Process-design and scale-up reasoning

Process design cannot be reduced to selecting plausible parameter values. Pharmaceutical digital twins may combine mechanistic, empirical, discrete-event, and process-data models to represent material and equipment behaviour [19]. Such hybrid modelling illustrates why linguistic reasoning alone cannot establish process state or scale transfer.

Machine learning may accelerate candidate prioritization for polymeric long-acting injectables by relating formulation attributes to observed outcomes [20]. Nevertheless, candidate ranking does not establish injectability, release robustness, stability, manufacturability, tolerability, or suitability for a specific product.

Continuous tablet compaction further requires measured process variables, control logic, equipment-specific behaviour, and implementation verification [21]. A generative system may organize process knowledge or suggest questions, but it should not autonomously issue control actions or authorize operating conditions.

The proposed process gate requires material properties, unit operations, equipment, scale, process parameters, critical quality attributes, control strategy, and disturbance context. Missing or contradictory information triggers rejection or escalation. Permission to test a process hypothesis remains distinct from approval to scale, manufacture, or release a product.

Uncertainty, traceability, and evidence warrants

Uncertainty should be treated as a property of the claim and decision context. Machine-assisted decisions require uncertainty assessment because apparently precise outputs may conceal limited data coverage, model instability, or unfamiliar inputs [22]. No single confidence value can represent all these sources.

Uncertainty communication should help users decide whether to accept, test, revise, or escalate a claim [23]. The architecture therefore separates uncertainty arising from missing evidence, conflicting sources, model extrapolation, unresolved mechanisms, process variability, and unknown product-context applicability.

Explainability is also purpose-dependent. An explanation may help a scientist inspect inputs or reasoning, but it cannot independently establish factual correctness, causality, compatibility, or product validity [24]. Traceable reasoning and scientific evidence must therefore remain distinct requirements.

The proposed warrant classes are direct observation, experimentally supported inference, mechanistically supported inference, empirical prediction, literature-grounded hypothesis, model-generated conjecture, and unsupported assertion. **Table 2** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for testable propositions, evidence requirements, and validation criteria for formulationgpt is not a formulator.

Table 2. Testable propositions, evidence requirements, and validation criteria for FormulationGPT Is Not a Formulator.

Architecture layer or process stage	Core function	Information or material flow	Interaction with other components	Validation criterion	Potential failure mode	Human or experimental responsibility	Readiness boundary
Intended-use gate	Define allowable task	User, context, output, consequence	Controls downstream evidence burden	Uses are correctly classified in prospective cases	Ambiguous or expanded use	Governance owner approves scope	No use outside defined context
Provenance and grounding	Link claims to evidence	Source-to-claim record	Supplies reviewable inputs [13-15]	Material claims are traceable and applicable	Missing, obsolete, or irrelevant source	Data steward and scientist verify	Traceability is not scientific validation

Compatibility layer	Screen candidates under constraints	Drug, excipient, route, grade, concentration	Feeds experimental hypotheses [16-18]	Constraint violations are detected; predictions tested prospectively	Hidden interaction or domain shift	Formulator authorizes testing	No autonomous compatibility conclusion
Process and scale-up layer	Test process applicability	Materials, equipment, scale, state	Connects formulation to manufacturing [19, 21]	Recommendations remain valid under defined equipment and scale	Incomplete process state	Process expert designs verification	No autonomous scale-up authorization
Evidence-warrant layer	Classify claim basis	Observation, inference, prediction, uncertainty	Determines review and escalation [22-24]	Reviewers distinguish evidence classes reliably	Confidence mistaken for evidence	Scientist assigns and challenges warrant	No unqualified decision claim
Human review gate	Exercise accountable judgment	Evidence package and unresolved risks	Controls experiment authorization [25, 26]	Reviewers detect errors and can reject output	Automation bias or nominal oversight	Qualified reviewer owns action	Human presence is not sufficient alone
Monitoring layer	Detect relevant change	Data, model, prompts, materials, workflow	Triggers reassessment [27, 28]	Material changes are detected and evaluated	Silent drift	Named owner initiates change control	Prior qualification may not transfer
Product-decision boundary	Prevent unauthorized commitment	Reviewed experimental and manufacturing evidence	Separates assistance from authority	No model output bypasses accountable approval	Direct automation of consequential action	Authorized organization makes decision	Not regulator-approved or deployment-ready

Human review and experiment-authorization gates

Artificial intelligence may augment professional work by expanding information access and analytical support, but professional responsibility remains necessary where context, consequences, and unresolved evidence matter [25]. The proposed architecture therefore assigns authority to qualified reviewers rather than to the generative system.

Human involvement is not automatically protective. Human–computer collaboration varies with task design, user expertise, interface, and the way recommendations are presented [26]. Reviewers may accept persuasive output despite weak evidence, especially when time pressure or automation bias reduces independent assessment.

A valid review gate requires access to source evidence, relevant scientific competence, authority to reject or revise output, and responsibility for the resulting action. Review disposition must be recorded as accepted for testing, revised, escalated, rejected, or outside intended use.

Experiment authorization is a separate gate. Authorization requires a defined hypothesis, justified materials and conditions, an appropriate protocol, safety considerations, acceptance logic, and a named approver. Approval to conduct an experiment does not validate the generated hypothesis or authorize subsequent manufacturing or product decisions.

Monitoring, change control, and accountability

Qualification at one time point cannot guarantee stable future behaviour. Model calibration may deteriorate as the relationship between inputs and outcomes changes, even when model code remains unchanged [27]. Monitoring must therefore examine performance within the evolving formulation context.

Dataset shift may arise from new materials, suppliers, analytical practices, dosage forms, user behaviour, or development priorities [28]. Changes in retrieval sources, prompts, tools, model versions, or interfaces may also alter outputs without an obvious change in the nominal task.

The proposed change-control process records the changed element, affected intended uses, potential consequences, required reassessment, reviewer decision, and implementation date. Material changes may trigger renewed grounding tests, constraint tests, uncertainty assessment, experimental verification, or suspension of the affected use.

Accountability requires named ownership for data, models, scientific review, experiment authorization, exceptions, monitoring, and product decisions. Auditability does not itself allocate legal liability, but it reduces ambiguity about who examined the evidence and who authorized each consequential transition.

Limitations and implementation priorities

Technical performance alone is insufficient for operational impact. Integration, external evaluation, usability, workflow compatibility, and organizational readiness influence whether an AI system produces benefit or avoidable disruption [29]. These requirements must be evaluated within pharmaceutical development rather than assumed from another domain.

Decision-support systems may provide useful structure while also introducing alert fatigue, over-reliance, workflow mismatch, and new error pathways [30]. Formulation-specific implementation studies should therefore examine how scientists use, challenge, override, and document generated recommendations.

The architecture is limited by the completeness of its evidence base and by differences among dosage forms, product classes, organizations, and development stages. It does not define validated thresholds, mandatory organizational roles, universal warrant categories, or a regulator-recognized assurance standard.

Implementation priorities include benchmark tasks tied to intended use, prospective compatibility and stability studies, equipment-specific process evaluation, adversarial constraint testing, uncertainty-calibration studies, reviewer-performance assessment, drift monitoring, and controlled pilot workflows. The architecture should be evaluated as falsifiable governance propositions rather than adopted as a complete deployment manual.

CONCLUSION

FormulationGPT is not a formulator because generated content does not independently integrate product context, experimental evidence, manufacturing state, uncertainty, professional accountability, and authority over pharmaceutical consequences. The proposed architecture separates assistance from authorization through intended-use classification, provenance and grounding, formulation and process constraints, evidence warrants, human review, experiment gates, product-decision boundaries, and lifecycle accountability. Its value lies in making hidden transitions explicit and testable. The architecture remains an original conceptual synthesis requiring computational, experimental, manufacturing, human-factors, and implementation validation. It does not establish formulation suitability, clinical benefit, regulatory acceptance, universal applicability, or deployment readiness.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. Adv Drug Deliv Rev. 2021;175:113806. doi:10.1016/j.addr.2021.05.016
2. Bao Z, Bufton J, Hickman RJ, Aspuru-Guzik A, Bannigan P, Allen C. Revolutionizing drug formulation development: The increasing impact of machine learning. Adv Drug Deliv Rev. 2023;202:115108. doi:10.1016/j.addr.2023.115108
3. Dong J, Wu Z, Xu H, Ouyang D. FormulationAI: A novel web-based platform for drug formulation design driven by artificial intelligence. Brief Bioinform. 2024;25(1). doi:10.1093/bib/bbad419
4. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. ACM Comput Surv. 2023;55(12):248. doi:10.1145/3571730
5. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172-80. doi:10.1038/s41586-023-06291-2
6. Oniani D, Hilsman J, Peng Y, Poropatich RK, Pamplin JC, Legault GL, et al. Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare. NPJ Digit Med. 2023;6(1):225. doi:10.1038/s41746-023-00965-x
7. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368. doi:10.1136/bmj.m689

8. 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
9. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517-8. doi:10.1001/jama.2017.7797
10. 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
11. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368. doi:10.1136/bmj.l6927
12. 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
13. 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
14. Hogan A, Blomqvist E, Cochez M, d'Amato C, de Melo G, Gutierrez C, et al. Knowledge graphs. ACM Comput Surv. 2021;54(4):71. doi:10.1145/3447772
15. Amugongo LM, Mascheroni P, Brooks S, Doering S, Seidel J. Retrieval augmented generation for large language models in healthcare: A systematic review. PLOS Digit Health. 2025;4(6). doi:10.1371/journal.pdig.0000877
16. Hang NT, Long NT, Duy ND, Chien NN, Phuong NV. Towards safer and efficient formulations: Machine learning approaches to predict drug-excipient compatibility. Int J Pharm. 2024;653:123884. doi:10.1016/j.ijpharm.2024.123884
17. Patel S, Patel M, Kulkarni M, Patel MS. DE-INTERACT: A machine-learning-based predictive tool for the drug-excipient interaction study during product development—validation through paracetamol and vanillin as a case study. Int J Pharm. 2023;637:122839. doi:10.1016/j.ijpharm.2023.122839
18. Han R, Xiong H, Ye Z, Yang Y, Huang T, Jing Q, et al. Predicting physical stability of solid dispersions by machine learning techniques. J Control Release. 2019;311-312:16-25. doi:10.1016/j.jconrel.2019.08.030
19. Moreno-Benito M, Lee KT, Kaydanov D, Verrier HM, Blackwood DO, Doshi P. Digital twin of a continuous direct compression line for drug product and process design using a hybrid flowsheet modelling approach. Int J Pharm. 2022;628:122336. doi:10.1016/j.ijpharm.2022.122336
20. Bannigan P, Bao Z, Hickman RJ, Aldeghi M, Häse F, Aspuru-Guzik A, et al. Machine learning models to accelerate the design of polymeric long-acting injectables. Nat Commun. 2023;14(1):35. doi:10.1038/s41467-022-35343-w
21. Bhaskar A, Barros FN, Singh R. Development and implementation of an advanced model predictive control system into continuous pharmaceutical tablet compaction process. Int J Pharm. 2017;534(1-2):159-78. doi:10.1016/j.ijpharm.2017.10.003
22. 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
23. Kompa B, Snoek J, Beam AL. Second opinion needed: Communicating uncertainty in medical machine learning. NPJ Digit Med. 2021;4(1):4. doi:10.1038/s41746-020-00367-3
24. 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
25. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
26. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229-34. doi:10.1038/s41591-020-0942-0
27. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc. 2017;24(6):1052-61. doi:10.1093/jamia/ocx030
28. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021;385(3):283-6. doi:10.1056/NEJMc2104626
29. 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

30. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: Benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17. doi:10.1038/s41746-020-0221-y