Qeios

Peer Review

Review of: "Rethink Your Mental Model in the Age of Generative AI: A Triadic Framework for Human-AI Collaboration"

Samuel Bassetto[1]

1. Mathematics and Industrial Engineering, École Polytechnique de Montréal, Université de Montréal, Canada

The article is strong as a conceptual and integrative framework paper, particularly in clarity of the central idea, depth of discussion, and bibliographic breadth.

However, under a strict scientific evaluation grid emphasizing quantification, hypotheses, empirical testing, systematic literature review, reproducibility, and traceability, it performs weakly because it does not aim to provide experimental validation or a systematic review methodology.

The priority improvements should be:

(a) Explicit formulation of testable hypotheses, (b) Systematic literature review protocol, (c) Defined measurable indicators, (d) Empirical validation with baseline comparison, (e) Reproducibility and data transparency statements.

1. Introduction

The opening successfully motivates why generative AI breaks familiar expectations (non-determinism, “jagged” capability, drift), but it remains mostly qualitative and does not anchor the stakes with at least one quantified order of magnitude tied to a single primary indicator that the paper will track. To strengthen it, you should add one short, traceable quantitative anchor (with a source) and explicitly state a central research question in one sentence, followed by a brief statement of scope (what the paper covers and what it does not). A clean outline placed at the end of the introduction would also improve navigability and reduce the feeling of a broad, essay-like start.

doi.org/10.32388/VIZIBU

2. Conceptual Foundations

This section does important definitional work (mental models, anthropomorphism, theory of mind), but the conceptual perimeter is sometimes too expansive, and the reader can lose the "operational payoff" of each definition. The main improvement is to tighten the concept set to only what is necessary for the framework and to add a compact glossary-style table mapping each key term to (i) a one-line definition, (ii) a concrete example, and (iii) a common misuse. If you keep the special notation (e.g., marking anthropomorphic terms), you should justify it with a short empirical or at least pragmatic argument ("how exactly does this change decisions or reduce errors?") and show one mini-case where the notation prevents a specific misunderstanding.

3. System Layer: Understanding

The diagnosis of system-level properties (variance, sensitivity to prompts, hallucinations, drift) is strong, but it blends claims of different epistemic status: some are well-established empirically, others are plausible generalizations. The improvement is to explicitly label claims as "supported by evidence," "observed in practice," or "hypothesis," and to add a compact evidence map (effect > task setting > model family/version > measured outcome > limitations). A second improvement is to standardize what "failure" means (e.g., factuality errors, citation fabrication, instruction-following breaks) and propose minimal measurement routines so that the layer can be evaluated rather than only described.

4. Collaboration Layer: Experiencing

The section convincingly argues that human-AI interaction fails when people import the wrong teamwork assumptions, but it sometimes reads as normative guidance without enough structured linkage to specific mechanisms (overtrust, automation bias, persuasion effects, social cues). To improve it, you should introduce a small taxonomy of collaboration failure modes, each with a concrete behavioral signature and a corresponding mitigation (what to do, when to do it, and how to detect it worked). It would also benefit from contrasting conditions where collaboration succeeds (counterexamples), because without that, the reader gets an asymmetrical picture dominated by failure narratives.

5. Metacognitive Layer: Reflecting

The metacognitive angle is a **real contribution**, but it risks becoming a catch-all bucket for "good habits" unless you make it more **testable** and **bounded**. The main improvement is to specify observable

metacognitive behaviors (e.g., independent verification, deliberate adversarial prompting, uncertainty logging, stopping rules) and connect them to measurable outcomes (error detection rate, calibration, decision latency, downstream harm proxies). A second improvement is to address feasibility: which practices are realistic under time pressure, what training is required, and what organizational incentives or tooling are necessary to make them routine rather than aspirational.

6. Summary of the Triadic Framework

The synthesis helps integrate the layers, yet the framework would be more compelling if it produced falsifiable predictions rather than only explanatory coherence. A key improvement is to state two or three clear predictions (e.g., "if X metacognitive routine is applied, then Y error type decreases under Z conditions") and describe what evidence would disconfirm them. Also, the boundaries between layers could be made sharper with a simple decision tree or diagnostic checklist showing how a practitioner distinguishes a system-layer problem from a collaboration-layer problem in real situations.

7. Propositions for Hybrid Intelligence

The propositions **are practical and readable**, but they currently function like guidelines rather than a research-ready program because they lack minimal operational definitions, success criteria, and evaluation designs. To strengthen this section, each proposition should include a short template: objective, required inputs, recommended procedure, failure modes, and one or two measurable indicators. You should also add at least one worked example end-to-end (a "toy" but realistic workflow) that applies multiple propositions and shows how the user would document verification, uncertainty, and responsibility.

8. Limitations and Critical Reflection

It is good that limitations are acknowledged, but they **remain broad** and could be sharpened into concrete threats to validity and transfer conditions. The improvement is to separate limitations into categories (construct validity, external validity, temporal validity due to model updates, and implementation constraints), and for each, state what the limitation would look like in practice and how future work could address it. This section should also explicitly state what would count as "evidence against" the framework, which increases scientific credibility and reduces the impression of an unfalsifiable position.

9. Future Research Agenda

The agenda is rich, but it is still high-level and would benefit from prioritization and a clearer connection to the propositions. The main improvement is to list a small set of “first experiments” with concrete designs (e.g., baseline vs. intervention, before/after, with/without routines), core metrics (calibration, factuality, reproducibility of outputs, decision quality), and recommended reporting standards (model/version, prompts, context windows, verification steps). You should also explicitly address reproducibility in a drifting-model world by proposing versioning practices, artifact sharing norms, and how to report when findings are model-dependent.

10. Conclusion

The conclusion is coherent and avoids excessive overclaiming, but it could do more work as a synthesis that a reader can reuse immediately. The improvement is to restate the central question and contribution in one crisp sentence, then summarize the framework and propositions in a compact way that mirrors the promised structure, and end with two or three concrete takeaways that are directly actionable. Finally, the conclusion should clearly reiterate the empirical gap: what is currently supported by evidence versus what remains proposed, so the paper closes with a precise boundary between knowledge and agenda.

Please find below a 50-criterion revision of your article. I gave the score: 2 for excellent, 1 for to be revised, and 0 for absent or critical.

#	Criterion	Score	Level	Justification
1	Title is clear, short, and reflects the topic	2	Good	Clear, precise, and aligned with the article's content.
2	Abstract presents stakes, method, results (quantified)	1	Needs Improvement	Stakes and conceptual approach are present, but no quantified results.
3	Three key points (Highlights) are provided	0	Critical	No explicit "Highlights" section with three structured points.
4	Stakes are clearly quantified with source and link to main indicator	0	Critical	Stakes discussed conceptually, not quantified with a traceable indicator.
5	References support the work	2	Good	Extensive bibliography; arguments are supported by peer-reviewed sources.
6	Research problem clearly stated as a question	1	Needs Improvement	Problem is clear but not formulated as a precise, explicit research question.
7	Depth of the research question	2	Good	Framework transcends a single application domain.
8	Working method is addressed	1	Needs Improvement	Conceptual synthesis described, but no formal methodology section.
9	Methodology includes review, hypotheses, procedure, discussion	0	Critical	No hypotheses nor experimental procedure.
10	Methodology appropriate to research type	0	Critical	No formal hypotheses or empirical protocol specified.
11	Document outline presented at end of introduction	1	Needs Improvement	Structure described, but not clearly formalized at the end of the introduction.
12	Reliability of the literature review	1	Needs Improvement	Review is rich but not systematic; no explicit search protocol.
13	Writing form of literature review	2	Good	Logical progression and thematic consistency overall.

#	Criterion	Score	Level	Justification
14	Industrial/organizational practices addressed	1	Needs Improvement	Practical routines proposed, but real-world case coverage limited.
15	Conclusion linked clearly to research question	2	Good	Returns to objectives and limitations without overclaiming.
16	Presentation and analysis of citations	1	Needs Improvement	Citations numerous, but not always critically dissected individually.
17	Types of literature search described	0	Critical	No verifiable description of databases, keywords, dates.
18	Hypotheses and GAP identification	1	Needs Improvement	GAP identified conceptually; no formal testable hypotheses.
19	Dedicated paragraph on notations	2	Good	Terminology and symbolic clarification explicitly discussed.
20	Each hypothesis discussed	0	Critical	No formal hypotheses presented.
21	Central idea clearly explained	2	Good	Triadic framework explained coherently and structurally.
22	Figures/tables referenced and explained	2	Good	Properly numbered, referenced, and integrated in text.
23	Example illustrating the idea	1	Needs Improvement	Examples exist but no fully simplified “toy” demonstration.
24	Paragraphs introducing and concluding the idea	2	Good	Logical sectional progression.
25	Idea explicitly described	2	Good	Core framework and routines are clearly articulated.
26	Limitations and validity explicitly discussed	1	Needs Improvement	Conceptual limitations discussed, but no empirical validation.
27	Experimental plan with baseline and analysis algorithms	0	Critical	No empirical plan or baseline comparison.
28	Quantifiable experiment with indicators and instruments	0	Critical	No experiment or measurement system presented.

#	Criterion	Score	Level	Justification
29	Data accessibility	0	Critical	No dataset associated with the work.
30	Detailed discussion section	2	Good	Substantial conceptual discussion and implications.
31	Basic statistics provided	0	Critical	No statistical results presented.
32	Hypothesis testing conducted	0	Critical	No hypothesis testing.
33	Data validity and limitations questioned	2	Good	Strong reflexivity on risks, hallucinations, overconfidence.
34	Implications clearly discussed	2	Good	Practical and research implications clearly developed.
35	Each figure/table referenced and explained	2	Good	Adequately integrated into the reasoning.
36	Synthetic conclusion	2	Good	Concise and coherent synthesis of contributions and limits.
37	Acknowledgments, ethics, data access	1	Needs Improvement	Declarations present, but no structured ethics statement or data policy.
38	Author and AI contribution declaration	1	Needs Improvement	Author contributions not fully detailed; no explicit AI usage declaration.
39	Funding and conflict of interest	2	Good	Funding and competing interests disclosed.
40	Bibliography completeness and consistency	1	Needs Improvement	Large and credible, but no explicit verification protocol stated.

Overall estimation

#	Criterion	Score	Justification
41	Traceable quantified stakes	0	No quantified stakes linked to measurable indicator.
42	All numerical claims sourced	0	Not all numerical effects formally traceable.
43	Performance comparison vs literature	0	Framework not empirically benchmarked.
44	AI usage declared	0	No explicit AI usage declaration.
45	Reproducibility of test and analyses	0	No test or reproducible protocol provided.
46	Verification of references via reliable database	0	No explicit verification statement.
47	Coherence intro ↔ results ↔ discussion	0	No measurable indicators announced and tested.
48	Coherence of hypotheses	0	No formal hypotheses stated.
49	Ethical dimension explicitly addressed	0	Ethical considerations discussed conceptually but no formal compliance statement.
50	Data accessibility	0	No open or reproducible dataset.

Declarations

Potential competing interests: No potential competing interests to declare.