

Iterative Critique of Therapeutic Resistance Narratives

Abstract

Iterative Critique of Therapeutic Resistance Narratives addresses a narrow but consequential interpretive problem: a plausible sequence of mechanisms lacking direct linkage. The article develops a research agenda through critical comparison of published evidence. The focal evidence concerns why autoimmune activity can persist or recur when conventional inflammation suppression does not durably change underlying cellular and tissue states. A chain of individually plausible statements can contain an unsupported transition. The gap may lie between exposure and pathway, pathway and state, or state and outcome. Audit every transition and attach its actual support. Mark unmeasured links explicitly and prioritize a test that could break the chain. The proposed comparison concerns neutrophil states, metabolic persistence, treatment resistance; it would test whether the specific failure scenario is detected without erasing legitimate uncertainty. Implications are restricted to the evidence roles relevant to immune-state and tumor evidence synthesis, leaving broader efficacy and deployment claims unresolved.

The failure scenario

Neutrophil states and treatment resistance occupy different positions in the evidence chain. Their relationship is the starting point for examining reasoning gaps. Consider the possibility of a plausible sequence of mechanisms lacking direct linkage. This is a proposed stress scenario, not an observed event reported by the cited studies. Its purpose is to expose which evidence contract would be violated even if the final output looked plausible. For tissue contexts, clinical endpoints, and mechanistic claims, the relevant claim must identify its population or artifact class, its time window, and the action it is supposed to support. A failure at any of these levels may persist even after a local technical improvement. The analysis therefore treats published methods as bounded evidence and asks how their limitations interact with this particular decision. An operational analogy is useful only if it keeps those limitations visible.

A chain of individually plausible statements can contain an unsupported transition. The gap may lie between exposure and pathway, pathway and state, or state and outcome. Fluency and self-consistency cannot supply the missing observation. In this article, the relevant contrast is neutrophil states, metabolic persistence, treatment resistance. The proposed counterexample of a plausible sequence of mechanisms lacking direct linkage puts that contrast under pressure. Audit every transition and attach its actual support. Mark unmeasured links explicitly and prioritize a test that could break the chain. A favorable finding would show that the intended interpretation of treatment resistance remains defensible when neutrophil states is perturbed under the specified conditions. An unfavorable finding should narrow the claim rather than be discarded as an inconvenient case.

The transition at issue is specific to immune-state and tumor evidence synthesis. It concerns neutrophil states, metabolic persistence, treatment resistance, and the scenario of a plausible sequence of mechanisms lacking direct linkage shows why a locally correct result can still become a misleading integrated claim. The input record should retain enough information to decide which part of that scenario violates the intended contract. That includes the original source type, measurement or artifact status, and the reason the proposed interpretation exceeds its support. The evaluator would compare the system's response with a qualified judgment made without seeing that response. Disagreement should be classified rather than erased by rewriting the prompt. Fluency and self-consistency cannot supply the missing observation. A useful negative result would show that the interface cannot preserve this distinction under realistic use; a useful positive

result would demonstrate preservation on independent cases rather than merely on the example used to develop the rule. The primary source remains relevant through its particular contribution: The set-point framing asks whether durable remission requires changing a coupled metabolic state rather than only lowering an inflammatory marker. That contribution should be retained even if the integrated workflow fails. Conversely, successful integration would not remove the source's limitations or turn a proposed extension into an already observed result. This asymmetric interpretation protects both the literature and the user from an unjustified all-or-nothing verdict.

Reading the literature by evidence type

(Xiao Yu et al. 2026e) examines why autoimmune activity can persist or recur when conventional inflammation suppression does not durably change underlying cellular and tissue states in *Resetting Immunometabolic Set Points in Autoimmune Disease*. Its methodological basis is a conceptual review proposing immunometabolic set points across immune and tissue compartments. The accessible evidence is specific: the review integrates cellular metabolism, nutrient/oxygen conditions, mitochondrial stress, and stromal metabolites as candidate mechanisms of persistence. The set-point framing asks whether durable remission requires changing a coupled metabolic state rather than only lowering an inflammatory marker. A unifying set point is a hypothesis, not an already validated clinical measurement or an established therapeutic protocol. For reasoning gaps, the important connection is the distinction between neutrophil states and treatment resistance. The source informs that distinction, but a stronger conclusion would require evidence at the point where one is used to stand for the other.

A useful boundary case is supplied by *Immune Niches in Brain Tumors: Glial-Immune Cell Interactions and Spatial Microdomains*, concerned with how spatial glial-immune interactions organize permissive and suppressive microdomains within brain tumors (Qian Guo et al. 2026). Its methodological basis is a review of astrocytes, microglia, oligodendrocyte-lineage cells, T cells, and myeloid cells across spatial tumor niches. The accessible evidence is specific: the abstract integrates cytokine, metabolic, hypoxic, and immunoediting mechanisms associated with immune suppression and exhaustion. It shifts attention from tumor-wide averages to interactions whose meaning depends on location and cell context. The review does not establish efficacy of a new intervention; co-location, state annotation, and functional interaction require different evidence. Reading the work through reasoning gaps changes the evaluation question: the issue is whether neutrophil states supports the intended interpretation of treatment resistance, rather than whether the reported method is impressive in its original task.

The relevant question in *Splenic Extramedullary Granulopoiesis and Neutrophil States in Cancer Immunity: Systemic Immunosuppression and Therapeutic Resistance* is how splenic extramedullary granulopoiesis and neutrophil heterogeneity contribute to systemic cancer immunosuppression and resistance (Yidan Zhang et al. 2026). Its methodological basis is a review organized around stress-associated splenic blood-cell production and reprogrammed neutrophil states. The accessible evidence is specific: the official Cancer Bioscience recent-articles listing confirms the review's scope and emphasizes sustained stress and systemic reprogramming; no new patient cohort is inferred. The framing connects extra-tumoral immune production with local cancer resistance questions. The accessible summary cannot establish specific quantitative contributions, universal cell-state functions, or clinical effects of proposed targeting strategies. The source supplies one part of the comparison involving neutrophil states, metabolic persistence, treatment resistance. Its evidence should retain its original unit and scope; the present synthesis proposes a connection without claiming that the paper directly tested this integrated workflow.

The relevant question in *The Prediction of Extended Hospital Length of Stay in Patients after Endoscopic Endonasal Transsphenoidal Surgery for the Resection of Non-Functioning Pituitary Adenomas:*

A Dual-Center Retrospective Analysis is which retrospective clinical features predict extended hospital stay after endoscopic endonasal surgery for nonfunctioning pituitary adenomas (Bibo Gao et al. 2026). Its methodological basis is a dual-center retrospective prediction analysis with a modeled extended-stay outcome and separate validation. The accessible evidence is specific: the accessible record describes a nomogram and a 100-patient validation sample; specific predictor effects and thresholds are not assumed. It treats hospitalization burden as an explicitly predicted outcome that can inform care-process evaluation. Length of stay includes organizational and discharge factors; it is not an interchangeable measure of tumor biology, surgical quality, or long-term recovery. The contribution becomes relevant to reasoning gaps when the workflow moves from neutrophil states to treatment resistance. That transition is the review's proposed analytical focus, not a new result attributed to the source.

The relevant question in AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought is whether structured self-critique can interrupt error propagation in language-model reasoning chains (Yinghao. “AutoCrit: A Meta-Reasoning Framework for Self-Critique Sang 2025). Its methodological basis is a reasoning agent, critique agent, and execution monitor provide iterative stepwise feedback and correction. The accessible evidence is specific: the abstract describes GSM8K, CSQA2, and ALFWorld evaluations, reported improvements over chain-of-thought baselines, and a fixed-point analysis of the correction process. It organizes critique as a monitored reasoning process rather than a single request to reconsider an answer. Shared model biases can correlate critic and reasoner errors, while extra iterations increase cost; benchmark gains do not establish clinical factuality. This source enters the present review as a methodological comparison for evidence processing, not as evidence about disease biology, patient outcomes, or clinical efficacy. The transferable question is whether neutrophil states remains correctly interpreted after a computational choice involving treatment resistance. Its task-specific validation cannot be borrowed to validate the clinical claim.

The sources should be compared through the roles they play, not merged into a single performance claim. Resetting Immunometabolic Set Points in Autoimmune Disease addresses why autoimmune activity can persist or recur when conventional inflammation suppression does not durably change underlying cellular and tissue states (Xiao Yu et al. 2026e). By contrast, AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought addresses whether structured self-critique can interrupt error propagation in language-model reasoning chains (Yinghao. “AutoCrit: A Meta-Reasoning Framework for Self-Critique Sang 2025). That difference matters for reasoning gaps: the evidence concerning neutrophil states is not automatically evidence concerning treatment resistance. A successful software operation, a plausible mechanistic account, and a validated outcome are different achievements. Their integration may improve traceability, but it cannot create an unmeasured link. A coherent review should identify where the sources complement one another and where their scopes remain incompatible. Here the proposed failure scenario is a plausible sequence of mechanisms lacking direct linkage. The analytical value of that scenario is that it can occur without requiring any cited paper to be wrong: an otherwise valid local finding can be misused at a later interface. Preventing that misuse requires evidence about the interface itself, not simply a larger collection of local results.

Methods that constrain the claim

Causal inference requires a clearly specified intervention, time order, and exchangeability assumptions (Hernán and Robins 2020). A prediction can be accurate without identifying what an intervention would change. A useful causal analysis begins by defining the hypothetical comparison, eligibility, and time zero. Its assumptions concern the process generating exposure and observation, not merely the flexibility of a fitted model. In this review, causal vocabulary is appropriate only when that comparison can be specified and defended. This is methodological support for examining reasoning gaps, not a claim that the source studied immune-state and tumor evidence synthesis. A separate application-specific test is required at the interface described in this review.

Epidemiologic design makes confounding, selection, and measurement error part of interpretation (Rothman, Greenland, and Lash 2008). Confounding, selection, and measurement error are distinct problems and need not move an estimate in the same direction. An analysis should describe them before deciding which adjustment is adequate. A larger dataset can reduce random error while leaving systematic error almost untouched. Precision should therefore be interpreted alongside the quality of the design and the eligibility of its population. This is methodological support for examining reasoning gaps, not a claim that the source studied immune-state and tumor evidence synthesis. A separate application-specific test is required at the interface described in this review.

False-discovery control is central when many molecular features are screened (Benjamini and Hochberg 1995). Multiplicity arises from the family of questions actually searched. Controlling false discoveries within a named family does not automatically cover later tissue filters, alternative endpoints, or narrative choices. The proposed audit should retain exploratory branches so the family is visible. Independent confirmation then addresses a different question from whether a discovery threshold was correctly applied during screening. The connection to iterative critique of therapeutic resistance narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Medical machine learning must fit data provenance, care pathways, and prospective evaluation (Rajkomar, Dean, and Kohane 2019). The clinical relevance of a model depends on how its output changes a care process. Development data can contain information that becomes available only after the intended decision. A review must identify that availability boundary before comparing algorithms. Prospective use also changes who is measured and who is treated, so evaluation of a frozen historical dataset cannot establish the entire future workflow. The connection to iterative critique of therapeutic resistance narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Responsible health AI connects prediction with workflow, monitoring, and patient impact (Wiens et al. 2019). The point at which a model reaches practice is not the end of evaluation. New populations, changing measurements, and feedback can alter failure modes. A responsible deployment needs a route for recognizing deterioration and a person authorized to respond. Monitoring should be linked to feasible corrective actions; an alert that cannot change the workflow provides accountability only in a very limited sense. For the comparison involving neutrophil states, metabolic persistence, treatment resistance, the source supplies an analytical obligation. The review's proposed application remains a hypothesis, with the deliberately difficult scenario serving as a check on scope inflation.

Calibration separates discrimination from confidence used in clinical escalation (Guo et al. 2017). Discrimination and calibration deserve separate evaluation. A score can order cases well yet systematically overstate its certainty. Aggregate calibration can also hide local miscalibration near the threshold where

action changes. An audit should retain predicted probabilities and observed outcomes rather than report only correct classifications, and it should evaluate calibration on data not reused to choose the correction. The connection to iterative critique of therapeutic resistance narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Uncertainty can deteriorate under the site and cohort shifts common in biomedicine (Ovadia et al. 2019). Uncertainty is most valuable when a system moves beyond familiar conditions, yet those conditions can also weaken its uncertainty estimates. A warning mechanism must therefore be evaluated under shift rather than assumed to generalize because it works on ordinary held-out data. Different shifts can affect ranking, calibration, and error concentration differently. No single in-distribution curve demonstrates protection against them all. The connection to iterative critique of therapeutic resistance narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Selective prediction formalizes abstention when evidence does not support action (Geifman and El-Yaniv 2017). Abstention creates a tradeoff between coverage and residual error. Comparing systems without matching coverage can make one look safer merely because it leaves more work unfinished. Deferred cases also consume human or computational resources. The evaluation should show the risk-coverage relation and account for the fate of excluded cases, rather than treating them as though they have ceased to matter. The connection to iterative critique of therapeutic resistance narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Feature attribution can guide audit but does not establish biological causality (Lundberg and Lee 2017). Attribution decomposes a model output according to a defined explanatory convention. That decomposition can reveal dependence on suspicious features without establishing why the outcome occurs in the world. Reference choices and feature dependence can affect interpretation. A useful audit should compare explanations with known model behavior and should not transform a feature contribution into a treatment or mechanism claim. The connection to iterative critique of therapeutic resistance narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Health-AI governance requires accountability, transparency, inclusion, and protection of autonomy (World Health Organization 2021). Health-AI governance concerns people affected by a system as well as the organization operating it. Transparency without a route to challenge an output is incomplete. Autonomy and inclusion require attention to who can obtain follow-up, who is represented in evaluation, and who bears false alarms. These obligations remain even when an assistance tool is described as advisory rather than decisive. For the comparison involving neutrophil states, metabolic persistence, treatment resistance, the source supplies an analytical obligation. The review's proposed application remains a hypothesis, with the deliberately difficult scenario serving as a check on scope inflation.

Research design and stopping conditions

The question left open by the literature is whether the intended distinction survives independent checking. Audit every transition and attach its actual support. Mark unmeasured links explicitly and prioritize a test that could break the chain. These procedures are proposed, not completed. The case set would include the scenario of a plausible sequence of mechanisms lacking direct linkage, with its validity judged independently of the system's response. The relevant comparison axes are neutrophil states, metabolic persistence, treatment resistance.

An evidence-role protocol would map each publication to its underlying observations and its permissible conclusions. The test cases would include reviews of the same primary study, correspondence without new data, and observations at incompatible scales. Reviewers would identify scope inflation before seeing the automated summary. An integrated system would be credited for preserving a missing link, not rewarded for making every source fit one complete narrative.

For the immune review, proposed mechanisms should be tested in their stated compartments and disease settings. Spatial association, state annotation, and intervention response would be judged separately. A useful result could expose a missing bridge in the integrated account. It would not need to confirm every part of the preferred narrative to improve scientific understanding.

What remains unverified

The cited sources still have design-specific limits even where their abstracts and methods are accessible. This review is a bounded conceptual synthesis rather than a systematic review with an exhaustive search protocol. It cannot establish the effectiveness of the proposed integrated workflow. Its specific contribution is to articulate how reasoning gaps changes the interpretation of neutrophil states, metabolic persistence, treatment resistance. Technical sources support computational comparisons only; they do not validate biological mechanisms or patient benefit. Reviews and correspondence retain their publication roles and are not counted as new experiments. Proposed stress cases are illustrative, not observed failures attributed to the papers. Future work should therefore test the weakest inference described here, preserve contrary results, and disclose the cost of checking rather than treating additional automation as intrinsically beneficial.

Conclusion

The counterexample of a plausible sequence of mechanisms lacking direct linkage shows why local success and integrated validity must be evaluated separately. Audit every transition and attach its actual support. Mark unmeasured links explicitly and prioritize a test that could break the chain. The central methodological contribution is specific: The set-point framing asks whether durable remission requires changing a coupled metabolic state rather than only lowering an inflammatory marker. Extending it to treatment resistance is an additional hypothesis, not a result supplied by the citation. A trustworthy workflow would preserve the evidence source, the assumptions of the transition, and the right to stop when support is insufficient. The proposed evaluation agenda gives those obligations observable tests while withholding claims of completed validation.

References

1. Yu, Xiao, et al. "Resetting Immunometabolic Set Points in Autoimmune Disease." *Journal of Translational Autoimmunity*, vol. 13, 2026, <https://doi.org/10.1016/j.jtauto.2026.100391>.
2. Guo, Qian, et al. "Immune Niches in Brain Tumors: Glial-Immune Cell Interactions and Spatial Microdomains." *Acta Neuropathologica Communications*, vol. 14, 2026, <https://doi.org/10.1186/s40478-026-02373-0>.
3. Zhang, Yidan, et al. "Splenic Extramedullary Granulopoiesis and Neutrophil States in Cancer Immunity: Systemic Immunosuppression and Therapeutic Resistance." *Cancer Bioscience*, 2026, <https://doi.org/10.1016/j.cbs.2026.100009>.
4. Gao, Bibo, et al. "The Prediction of Extended Hospital Length of Stay in Patients after Endoscopic Endonasal Transsphenoidal Surgery for the Resection of Non-Functioning Pituitary Adenomas: A Dual-Center Retrospective Analysis." *Cancers*, vol. 18, no. 10, 2026, <https://doi.org/10.3390/cancers18101582>.
5. Sang, Yinghao. "AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought." *2025 6th International Conference on Machine Learning and Computer Application (ICMLCA)*, IEEE, 2025, pp. 1177-80, <https://doi.org/10.1109/icmlca66850.2025.11336788>.
6. Hernán, Miguel A., and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.
7. Rothman, Kenneth J., Sander Greenland, and Timothy L. Lash. *Modern Epidemiology*. 3rd ed., Lippincott Williams & Wilkins, 2008.
8. Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B*, vol. 57, no. 1, 1995, pp. 289-300.
9. Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. "Machine Learning in Medicine." *New England Journal of Medicine*, vol. 380, 2019, pp. 1347-58.
10. Wiens, Jenna, et al. "Do No Harm: A Roadmap for Responsible Machine Learning for Health Care." *Nature Medicine*, vol. 25, 2019, pp. 1337-40.
11. Guo, Chuan, et al. "On Calibration of Modern Neural Networks." *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp. 1321-30.
12. Ovadia, Yaniv, et al. "Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty under Dataset Shift." *Advances in Neural Information Processing Systems*, vol. 32, 2019.
13. Geifman, Yonatan, and Ran El-Yaniv. "Selective Classification for Deep Neural Networks." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
14. Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
15. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*. World Health Organization, 2021.