
Tracking Scientific Information Loss in LLM-Assisted Biomanufacturing Process Development

Sebastián A. Cruz Romero*
Capicú Technologies
scruzromero@capicu.ai

Abstract

Scientific assistants must preserve evidence as it moves from experimental design to comparison, decision, and operating instructions. We introduce BIOPROCESSLINEAGE, an evaluation with 12 linked workflows, 78 source-backed facts, and 109 fact-stage requirements from 11 openly licensed publications. Across nine hosted open-weight model routes, conventional screens are strong, yet direct outputs retain only 40.1% of complete facts and 7.2% of multi-stage facts consistently, despite 95.6% exact-evidence precision. This separates grounding (support for included claims) from completeness. In the common eight-route matched comparison, model-created lineage and graph structures reduce retention, source-unit coverage certificates add 10.3 percentage points, and deterministic fact-slot compilation reaches 93.5%. The latter preserves content by design but exposes stage routing as the next bottleneck. Controlled source interventions support this diagnosis: removing a target evidence unit lowers target-term retention by 46.7 points, while restoring it recovers 44.5 points. The results motivate evaluating scientific systems as information pipelines and protecting source facts before generation.

1 Introduction

Biomanufacturing process development repeatedly transforms evidence. A study defines variables and ranges where its results support a comparison and decision leading to a decision that becomes an operating procedure. A value, comparator, or scope qualifier established early may be needed in every later artifact. Its omission can leave fluent text that no longer represents the supporting experiment.

Current evaluations are poorly aligned with this risk. End-task accuracy can reward the selected condition without its range or scope. Factuality metrics usually test support for generated claims rather than coverage of required source facts [Min et al., 2023, Wei et al., 2024]. Recall-oriented work addresses part of this gap [Liu et al., 2025, Jafari et al., 2026]; a linked scientific workflow adds stage-specific obligations and repeated use of the same fact.

We therefore ask whether broad task scores predict information retention, whether structured generation protects facts, and whether source evidence controls downstream artifacts. BIOPROCESSLINEAGE tests these questions with a hidden source-fact inventory, deterministic scorers, matched interventions, and nine hosted open-weight model routes. It measures where information is lost without using an LLM judge.

The result is a consistent metric gap: task performance, valid structure, and exact quotations do not imply conservation. Direct outputs preserve only 40.1% of strict fact-stage requirements. Structures created by the model inherit its omissions; exhaustive source accounting helps; deterministic copying protects coverage but over-routes facts. Controlled source interventions show that relevant evidence changes outputs, locating the main measured bottleneck between evidence availability and reliable

*Corresponding author.

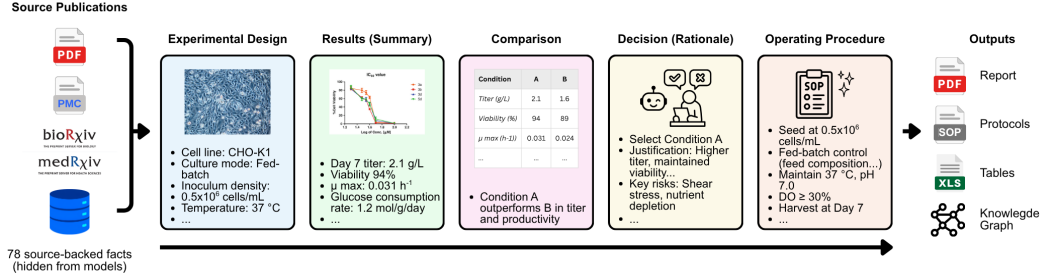


Figure 1: **Linked biomanufacturing information flow.** Evidence from source publications is transformed through experimental design, results, comparison, decision rationale, operating procedures, and structured outputs. A hidden inventory of 78 source-backed facts and 109 fact-stage requirements tracks whether values, scope, and rationale survive each transformation.

propagation. Our contributions are the fact-stage benchmark, a matched comparison of seven preservation designs, and an artifact-level causal test of this bottleneck.

2 Related work

Atomic factuality methods test support for generated claims [Min et al., 2023, Wei et al., 2024]; reference-fact and importance-aware methods add coverage [Liu et al., 2025, Jafari et al., 2026]. We extend this precision-recall distinction to stage-specific requirements: a fact is preserved only where the workflow needs it, and multi-stage facts must remain complete end to end.

Bioprocess agents combine retrieval, decomposition, planning, verification, or typed protocol compilation [You et al., 2026, Liu et al., 2026, Song et al., 2026]. We study the information boundary before those controls: whether the source facts reach every artifact and intermediate representation they depend on. Our method names denote controlled design patterns, not full reimplementations of these systems.

3 Problem setting and benchmark

3.1 Fact lineage

A workflow $w = (s, o, T, F)$ contains a bounded source passage s , an objective o , requested stages T , and hidden reference facts F . Each fact f has an exact evidence quotation, required terms K_f , optional scope terms, and a set of stages $T_f \subseteq T$ where the fact is needed. The model receives s , o , and the stage contracts, but never F or T_f .

For generated stage text y_t , strict fact retention is

$$R_{\text{fact}} = \frac{1}{\sum_f |T_f|} \sum_{f \in F} \sum_{t \in T_f} \mathbb{1}[\forall k \in K_f, k \in y_t]. \quad (1)$$

Term retention gives partial credit to individual required terms. Cross-stage consistency credits a multi-stage fact only if all its terms appear in every required stage. Scope preservation applies the same check to cell line, condition, material, or process qualifiers. Fixed normalization handles case, spacing, and punctuation. Exact-evidence precision measures submitted quotations found verbatim in the source; numerical support, schema validity, and stage completion measure separate error modes. No scorer uses semantic embeddings or an LLM judge.

Figure 1 summarizes the linked information flow; Figure 2 distinguishes early omission from loss during generation or downstream propagation.

3.2 Workflow construction

BIOPROCESSLINEAGE draws 12 workflows from 11 peer-reviewed, openly licensed biomanufacturing articles published from 2017–2023 [Böhl et al., 2021, Pan et al., 2017, Torres et al., 2021,

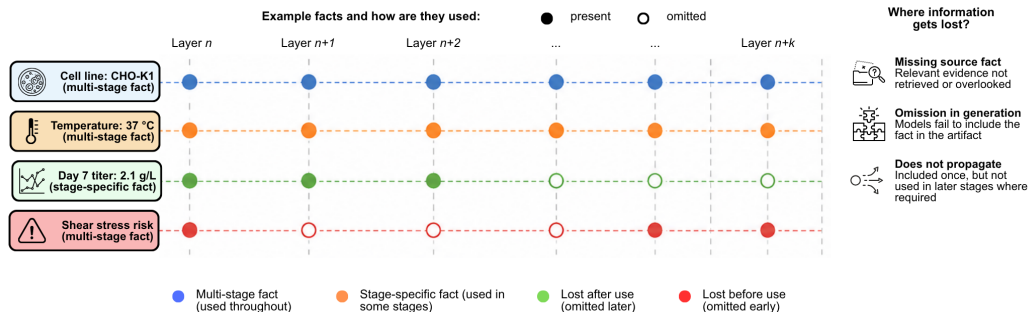


Figure 2: **Observable paths of scientific-information loss.** A required fact may be absent before generation, omitted while an artifact is generated, or present once but not propagated to a later stage. Filled and open markers denote presence and omission across workflow layers; colors distinguish multi-stage facts, stage-specific facts, and early or late loss. This representation motivates strict fact–stage retention and cross-stage consistency.

Table 1: **BIOPROCESSLINEAGE contents.** The 12 workflows contain 32 requested stage instances across 11 source groups, 78 reference facts, and 109 fact–stage requirements. Examples are shortened source-backed passages; bold text marks the content relevant to each stage.

Workflow stage	n	Information tracked	Example passage
Experimental design	8	Factors, ranges, controls, and responses	"A <i>three-factor, three-level CCD</i> used 19 experiments."
CPP/CQA identification	1	Identities of process parameters and quality-attributes	" <i>Methotrexate concentration</i> was a critical parameter."
Condition comparison	6	Conditions and quantitative outcomes	"The shift increased Titermax by 54% in BC and 111% in FB."
Process decision	11	Selected conditions, rationale, and bounds	"At 32 °C, viability decline was delayed by 2 days in BC and 5 days in FB."
Standardization	6	Setpoints, sequence, and acceptance criteria	"Flasks were incubated at 37 °C and 220 rpm for 90 h."

Marschall et al., 2023, Zhang et al., 2023, Behravan and Hashemi, 2021, Agbogbo et al., 2020, Gotsmy et al., 2023, Liu et al., 2020, Ghorbani Aghdam et al., 2019, Sakhnini et al., 2019]. It covers five linked stages; 30 facts are required in multiple stages. Fact-slot scoring defines 107 source-unit–stage gold links; proposition routing uses 314 candidates, including 127 required edges. Each record includes its exact excerpt, license, URL, and excerpt hash. Validation checks evidence membership, source integrity, and fact–stage consistency. The inventory was authored independently of model outputs. Table 1 summarizes the workflow contents; stage counts overlap because one workflow can produce several linked artifacts.

4 Experimental design

4.1 Model panel and common controls

The panel comprises Kimi-K3, GLM-5.2, GPT-OSS-120B, Qwen3.5-397B-A17B, DeepSeek-V4-Flash, MiniMax-M3, Gemma-4-31B-IT, Qwen3.5-9B, and Qwen3.6-35B-A3B, routed through Together, Groq, or DeepInfra. We use temperature 0, seed 20260828, low or disabled reasoning output, and JSON-object mode when exposed. Predictions are fingerprinted and saved record-by-record.

Output budgets are 2,304 tokens for lineage and certificate calls, at most 1,024 for graph calls, 96 per fact-slot routing decision, and 512 per causal artifact. Generated lineage stages are capped at 180 words and causal stages at 70. Because methods differ in calls and deterministic post-processing, we compare complete systems rather than equal-compute prompts. Each panel reports its exact available-route denominator ($n = 8$ or 9) without imputation.

4.2 Capability screens

Three bounded screens test whether conventional capability anticipates preservation. We use the 80 common text items from BioManufacturingBench v1.1 as a descriptive screen; its full pilot contained 41% zero-variance items and 15.3% negative discrimination estimates, so rank differences are not treated as a validated leaderboard. Four SOP cases measure required and critical terms, strict requirement coverage, target-text overlap, JSON validity, and numerical literal support against curated target text. Five process cases measure factor recall, CPP/CQA element recall, condition-detail recall, decision critical-term recall, and standardization term recall. These screens provide context for the linked-workflow measurements.

4.3 Preservation interventions

Every generative method receives the same workflow objective, source, requested stages, scope-preservation rule, and prohibition on unsupported claims. The interventions test which controls reduce the loss patterns in Figure 2.

Generation and verification. *Direct* generates every stage with short exact quotations. *Hierarchical decomposition* first lists up to 12 facts and constraints within the same response. *Generate–repair* adds a second call with deterministic feedback for malformed JSON, missing stages, unsupported numbers, and inexact quotations; hidden reference facts are never revealed.

Model-created structure and source accounting. *Lineage* asks the model to create shared fact identifiers cited by each stage. *Coverage certificates* instead segment the visible source into numbered units that must be accounted for and linked to stages. *Dependency graphs* ask the model to construct source-unit nodes and dependencies before stage generation. Verified lineage and certificate variants receive one source-visible repair call.

Recall-protected compilation. *Fact slots* present source-exact units as FactIDs. The model selects IDs for each stage, after which software copies the corresponding text verbatim. Invalid routes select all units, and globally omitted units are appended to every stage. This design protects recall and makes routing precision measurable.

One-pass lineage and certificates are paired with direct generation; verified lineage, verified certificates, graphs, and fact slots are paired with generate–repair. Deltas are computed within route before aggregation.

4.4 Causal source-evidence intervention

We select one isolated target per workflow whose evidence maps to one source unit and no other annotated fact. Ranking prioritizes numerical, multi-stage, term-rich, and critical facts, yielding 12 targets. Each is tested with the original source (*clean*), the target unit removed (*ablated*), the target restored as a provenance-labeled packet (*restored*), or an unrelated fact in the same packet format (*sham*). Necessity is clean–ablated retention, restoration is restored–ablated, and specificity is restored–sham. These contrasts identify effects of controlled input evidence on generated artifacts.

4.5 Aggregation and uncertainty

Metrics are averaged equally across workflows within each route, then equally across available routes. Intervention effects are paired within route. We report exact denominators, observed ranges, and descriptive Spearman correlations; no population inference or formal hypothesis test is intended.

5 Results

5.1 Broad capability does not predict strict preservation

Figure 3 gives the common eight-route cross-experiment view: mean common-text benchmark and isolated process scores are 80.2% and 85.5%, respectively. Across all nine routes, direct strict fact retention is 40.1% (range 18.5–65.9%) and cross-stage consistency is 7.2%. The benchmark–retention

Table 2: **Measured information-preservation outcomes.** Percentage-point (pp) changes use the matched baseline within each model route.

Measurement	Observed result	<i>n</i>
Grounding vs. completeness	95.0% exact-evidence precision; 38.6% strict fact retention	8
Cross-stage conservation	6.9% of multi-stage facts complete throughout	8
Coverage certificate	+10.3 pp retention relative to direct generation	8
Fact-slot compiler	93.5% retention; +53.4 pp relative to generate–repair	8
Fact-slot routing	53.1% precision at 97.3% recall	8
Evidence intervention	−46.7 pp after ablation; +44.5 pp after restoration	9

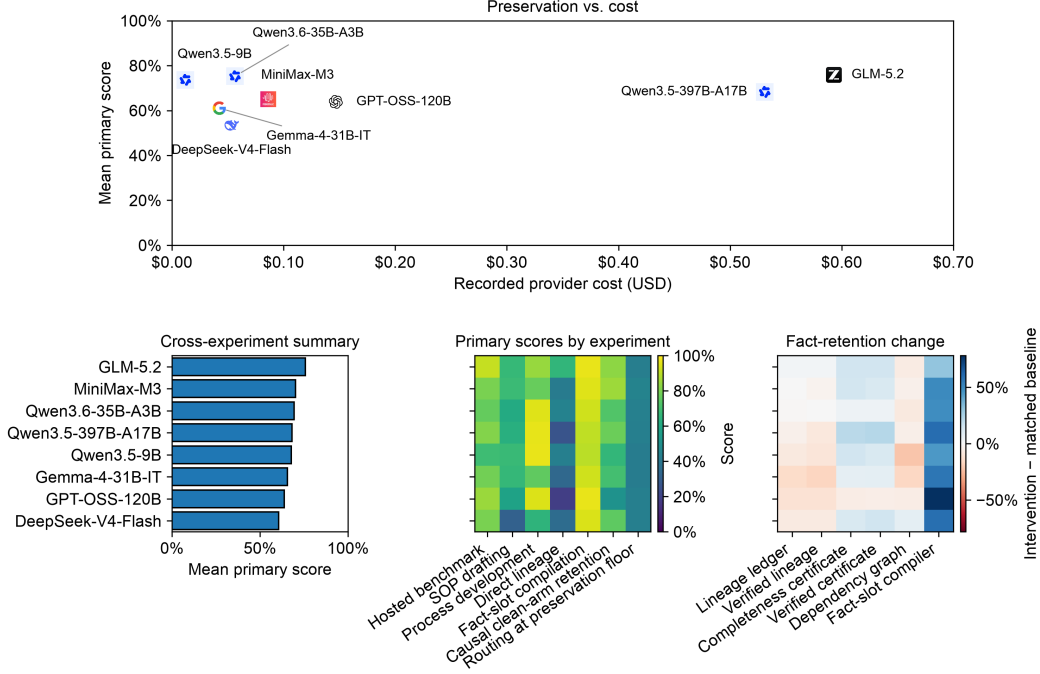


Figure 3: **Cross-model scientific-information preservation.** Recorded cost, the seven-endpoint descriptive summary, experiment weights, and matched retention changes for the common eight-route subset. The summary equally weights heterogeneous endpoints and the cost is runner-estimated; both are descriptive.

rank correlation is only $\rho = 0.126$ ($n = 9$). GPT-OSS-120B illustrates the mismatch: 85.0% on the common-text screen and 94.6% across process metrics, but 18.5% strict retention.

The common eight-route subset in Figure 3 averages 67.7% on the unweighted seven-endpoint summary (range 60.6–75.9%). Recorded preservation-run cost ranges from \$0.022 to \$0.593; these configured token-rate estimates provide run context rather than a normalized cost-effectiveness comparison. Table 2 summarizes the main measured preservation outcomes; Figures 3–5 show their model-level structure.

5.2 Grounding and completeness measure different behavior

Figure 4 resolves the metric gap. Across nine direct routes, term retention is 67.0%, complete-fact retention 40.1%, critical-fact retention 39.1%, and cross-stage consistency 7.2%. In the common eight-route set, direct outputs attain 95.0% exact-evidence precision, 99.1% supported numerical literals, and 99.0% stage completion. Grounding is therefore strong for included content while required content remains incomplete.

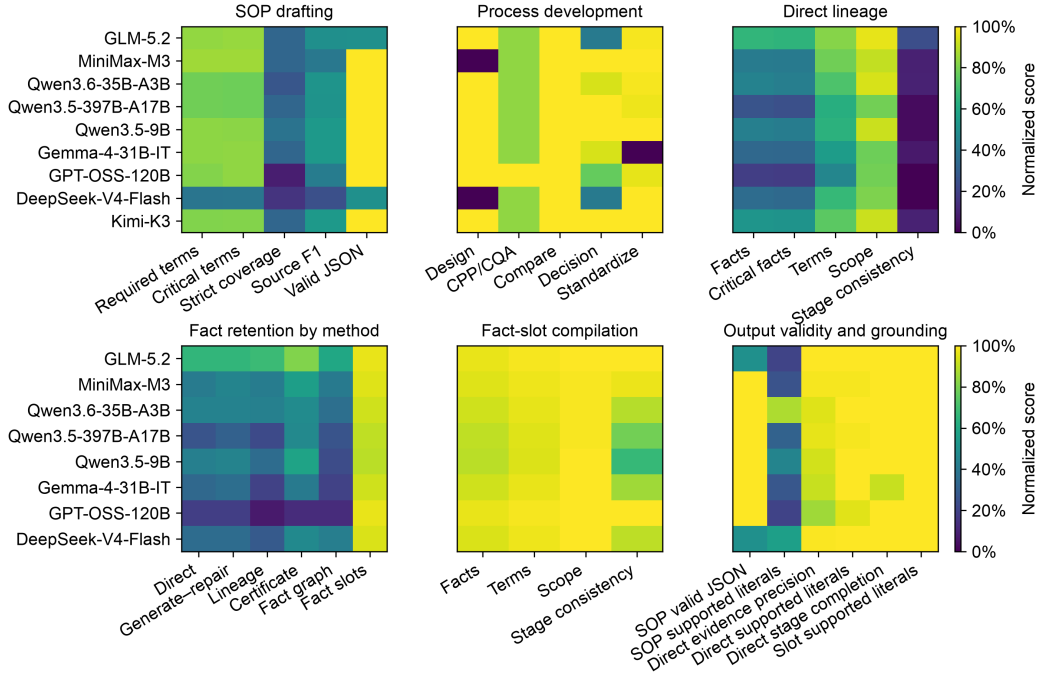


Figure 4: **Detailed capability, preservation, and reliability metrics.** The top row uses nine routes for SOP, process, and direct-lineage metrics; the bottom uses the common eight-route subset for method, fact-slot, and reliability metrics. Each row shares a route axis and 0–100% scale. SOP “Source F1” uses curated target text; direct evidence precision uses the source passage.

SOP outputs show the same partial-credit pattern: required- and critical-term recall average 77.1% and 77.3%, but strict requirement coverage is 28.1% ($n = 9$). Process metrics range from 77.8% to 100%; condition comparison is saturated across all nine displayed routes. These isolated scores do not expose whether every required detail survives a linked workflow.

5.3 Source accounting protects facts; self-generated structure does not

Figure 5 compares matched effects in the common eight-route set. Lineage changes retention by -5.2 points, verified lineage by -6.6 , and dependency graphs by -7.4 ; six, six, and seven routes respectively show decreases. A representation built from incomplete extraction preserves the omission. Coverage certificates reverse the direction: one-pass certificates add 10.3 points over direct generation and verified certificates add 10.4 over generate-repair, with increases in seven routes. Exhaustive source accounting helps, although certificate retention remains 48.9%.

The fact-slot compiler produces the largest increase: 93.5% mean strict retention and a 53.4-point gain over generate-repair, increasing all eight routes. Term retention reaches 96.9%, scope preservation 99.6%, and every numerical literal is source-supported. This is a compiler result: software copies selected units and appends globally omitted units to every stage. Consequently, routing recall is 97.3% but precision is 53.1%. A multisignal proposition router retains the coverage floors at 43.1% precision. The system can protect content; its next design objective is precise stage assignment.

5.4 Evidence interventions locate the propagation bottleneck

The causal panels in Figure 5 isolate one target source unit per workflow. Across nine routes, removal lowers target-term retention by 46.7 points and complete-target retention by 36.1. Restoration recovers 44.5 term points and 32.9 complete-fact points; relative to sham restoration, the target packet adds 46.1 and 32.9 points. Every route shows positive effects. Relevant evidence therefore controls artifact content, while direct generation still omits most complete fact-stage requirements.

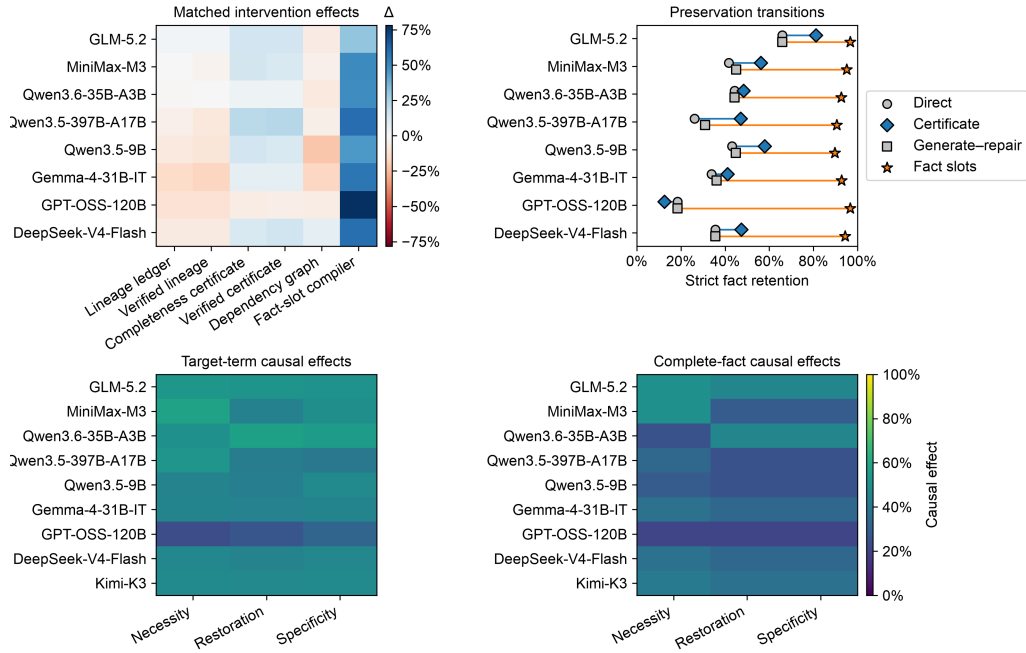


Figure 5: **Preservation mechanisms and artifact-level causal effects.** Top: matched retention effects and transitions for eight routes. Bottom: necessity (clean—ablated), restoration (restored—ablated), and specificity (restored—sham) for terms and complete facts across nine routes. The interventions identify effects on generated artifacts, not internal model mechanisms.

The measured bottleneck lies in selecting that evidence, assigning it to each dependent stage, and preserving all values and qualifiers.

6 Scope and future evaluation

This study targets literal conservation across 12 workflows from 11 source groups. Exact-term scoring makes numerical values, ranges, and scope auditable; future evaluation can pair it with expert adjudication and paraphrase-aware semantic scoring, while expanding CPP/CQA and modality coverage. The capability screens serve as context rather than a leaderboard; larger calibrated item banks can test how preservation relates to domain capability.

The comparisons describe model-provider routes under one fixed generation setting and evaluate complete workflows rather than equal-compute prompts. Pinned local checkpoints, repeated generations, provider replication, and factorial control of calls and output length are natural extensions. The causal experiment isolates effects of source packets on generated artifacts; local models would additionally support hidden-state and token-probability analyses, and contamination probes can separate source use from memorization.

The intended use is evidence-preservation research, not autonomous process control. Process experts, quality review, and laboratory execution provide the next validation layer for any operational system.

7 Discussion and conclusion

The experiments support a simple mechanism. Standard generation optimizes a plausible answer, while conservation requires every needed fact to be selected and propagated. Verification can detect unsupported numbers or inexact quotes, but it cannot repair an omission without first representing

what is missing. Lineage ledgers and graphs therefore inherit extraction gaps; adding structure after selection does not recover absent content.

Exhaustive source accounting changes that sequence. Certificates require visible units to be considered before generation, and fact-slot compilation prevents rewriting from changing them. This raises retention from roughly 40% to 94%, while moderate routing precision identifies the remaining optimization target. A practical architecture is therefore recall-protected compilation followed by precision control and human review at the fact-stage boundary.

Scientific assistants should be evaluated as information pipelines, not only as answer generators. The source interventions show that models use relevant evidence; the linked benchmark shows that they do not consistently carry all of it forward. Measuring that gap makes preservation an explicit engineering objective.

Reproducibility statement

The evaluation configuration records model and provider routes, prompt and source hashes, deterministic scoring code, generation options, a fixed seed, and record-level fingerprints. Raw responses support deterministic re-scoring without regeneration.

References

- Frank K. Agbogbo, Phil Ramsey, Renija George, Jobin Joy, Shikha Srivastava, Mian Huang, and Jesse McCool. Upstream development of *Escherichia coli* fermentation process with PhoA promoter using design of experiments (DoE). *Journal of Industrial Microbiology & Biotechnology*, 47 (9–10):789–799, 2020. doi: 10.1007/s10295-020-02302-7. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC7658055/>.
- Aidin Behravan and Atieh Hashemi. RSM-based model to predict optimum fermentation conditions for soluble expression of the antibody fragment derived from 4D5MOC-B humanized MAb in SHuffle T7 *E. coli*. *Iranian Journal of Pharmaceutical Research*, 20(1):254–266, 2021. doi: 10.22037/ijpr.2020.114377.14822. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC8170757/>.
- Ole Jacob Böhl, Jana Schellenberg, Janina Bahnemann, Bernd Hitzmann, Thomas Scheper, and Dörte Solle. Implementation of QbD strategies in the inoculum expansion of a mAb production process. *Engineering in Life Sciences*, 21(3–4):196–207, 2021. doi: 10.1002/elsc.202000056. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC7923587/>.
- Atefeh Ghorbani Aghdam, Saeed Moradhaseli, Farnoush Jafari, Paria Motahari, Sepideh Samavat, Rasoul Mahboudi, and Shayan Maleknia. Therapeutic Fc fusion protein misfolding: A three-phasic cultivation experimental design. *PLOS ONE*, 14(1):e0210712, 2019. doi: 10.1371/journal.pone.0210712. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC6334962/>.
- Mathias Gotsmy, Florian Strobl, Florian Weiß, Petra Gruber, Barbara Kraus, Juergen Mairhofer, and Jürgen Zanghellini. Sulfate limitation increases specific plasmid DNA yield and productivity in *E. coli* fed-batch processes. *Microbial Cell Factories*, 22(1):242, 2023. doi: 10.1186/s12934-023-02248-2. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC10685491/>.
- Nazanin Jafari, James Allan, and Mohit Iyyer. Beyond precision: Importance-aware recall for factuality evaluation in long-form LLM generation, 2026. URL <https://arxiv.org/abs/2604.03141>.
- Wancang Liu, Haibo Xiang, Tao Zhang, Xu Pang, Jing Su, Hongyu Liu, Baiping Ma, and Liyan Yu. Development of a new high-cell density fermentation strategy for enhanced production of a fungus β -glucosidase in *Pichia pastoris*. *Frontiers in Microbiology*, 11:1988, 2020. doi: 10.3389/fmicb.2020.01988. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC7472535/>.
- Xin Liu, Lechen Zhang, Sheza Munir, Yiyang Gu, and Lu Wang. VeriFact: Enhancing long-form factuality evaluation with refined fact extraction and reference facts. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 17908–17925.

- Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.emnlp-main.905. URL <https://aclanthology.org/2025.emnlp-main.905/>.
- Yuyang Liu, Jingya Wang, Liuzhenghao Lv, and Yonghong Tian. BioProAgent: Neuro-symbolic grounding for constrained scientific planning. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 42764–42783. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.1981. URL <https://aclanthology.org/2026.acl-long.1981/>.
- Lukas Marschall, Chitti Babu Gottimukkala, Biswajit Kayal, Veerabhadra Madurai Veeraraghavan, Samir Kumar Mandal, Suman Bandyopadhyay, and Christoph Herwig. Temperature upshifts in mammalian cell culture: A suitable strategy for biosimilar monoclonal antibodies? *Bioengineering*, 10(10):1149, 2023. doi: 10.3390/bioengineering10101149. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC10603922/>.
- Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. FActScore: Fine-grained atomic evaluation of factual precision in long form text generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12076–12100. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.emnlp-main.741. URL <https://aclanthology.org/2023.emnlp-main.741/>.
- Xiao Pan, Mathieu Streefland, Ciska Dalm, René H. Wijffels, and Dirk E. Martens. Selection of chemically defined media for CHO cell fed-batch culture processes. *Cytotechnology*, 69(1):39–56, 2017. doi: 10.1007/s10616-016-0036-5. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC5264622/>.
- Laila I. Sakhnini, Anja K. Pedersen, Ileana Rodríguez León, Per J. Greisen, Jens Jacob Hansen, Malene B. Vester-Christensen, Leif Bülow, and Maria B. Dainiak. Optimizing selectivity of anion hydrophobic multimodal chromatography for purification of a single-chain variable fragment. *Engineering in Life Sciences*, 19(7):490–501, 2019. doi: 10.1002/elsc.201800207. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC6999250/>.
- Renjian Song, Yaokai Fu, Ziyang Zhao, Jigang Yu, Qing Yuan, and Chang-Ting Chen. Towards autonomous biology: Compiler-verified protocols as a foundation for real-world AI execution. *bioRxiv*, 2026. doi: 10.64898/2026.05.05.720956. URL <https://www.biorxiv.org/content/10.64898/2026.05.05.720956v1>. Preprint.
- Mauro Torres, Samia Akhtar, Edward A. McKenzie, and Alan J. Dickson. Temperature down-shift modifies expression of UPR-/ERAD-related genes and enhances production of a chimeric fusion protein in CHO cells. *Biotechnology Journal*, 16(2):e2000081, 2021. doi: 10.1002/biot.202000081. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC11475507/>.
- Jerry Wei, Chengrun Yang, Xinying Song, Yifeng Lu, Nathan Hu, Jie Huang, Dustin Tran, Daiyi Peng, Ruiibo Liu, Da Huang, Cosmo Du, and Quoc V. Le. Long-form factuality in large language models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/937ae0e83eb08d2cb8627fe1def8c751-Abstract.html.
- Xiaoyu You, Zhenhua Yu, Di Wu, Luyun Shi, Guan Wang, Zhixing Cao, Qingchao Jiang, Xuefeng Yan, Weimin Zhong, and Biao Huang. BioMeter and BioManner: Evaluating and enhancing reliability of large language models in bioprocess engineering. *Digital Chemical Engineering*, 20:100318, 2026. doi: 10.1016/j.dche.2026.100318. URL <https://doi.org/10.1016/j.dche.2026.100318>.
- Ying Zhang, Jingmin Hu, Qing Zhang, Dongbo Cai, Shouwen Chen, and Yonghong Wang. Enhancement of alkaline protease production in recombinant *Bacillus licheniformis* by response surface methodology. *Bioresources and Bioprocessing*, 10(1):27, 2023. doi: 10.1186/s40643-023-00641-8. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC10991860/>.