

FORSCH-PROV0459_pamphlet-de_v0.5_2026-10-06_werk

title: "Iteratio Simulata — Pamphlet"
subtitle: "Performative Iteration in LLM-Outputs erkennen"
author: "Thomas Schüller"
orcid: "0009-0003-9799-6747"
contact: "forschung@tschueller.com"
version: "Pamphlet v0.4"
date: "2026-10-06"
status: "Kurzfassung der Working Note B für externe Lektüre"
license: "CC BY 4.0"
kennung: "FORSCH-PROV0459"
slug: "iteratio-simulata"
register_hinweis: „Nummer provisorisch; finale Vergabe gegen das Werkregister“

Iteratio Simulata

Wenn ein einzelner LLM-Output die Struktur einer Forschungssitzung imitiert, ohne sie auszuführen.

Worum es geht

In LLM-Outputs taucht zunehmend ein Muster auf: Ein Prompt fordert "15 Iterationen", "harte Selektion", "kritischen Reviewer, der schwache Konzepte verwirft" — und das Modell liefert einen Text, der diese Form rhetorisch perfekt imitiert, ohne dass die Iteration tatsächlich stattfindet. Konzepte werden umbenannt, aber nicht eliminiert. Bewertungskriterien tauchen nach der Bewertung auf. Etablierte Forschung wird als Durchbruch reframet. Über aufeinanderfolgende Iterationsschritte werden Zahlen, Preise und Margen immer präziser, ohne dass eine einzige davon belegt wäre.

Das Phänomen ist **kein Faktenfehler** (die Einzelaussagen können stimmen) und **keine bloße Sycophancy** (der Ton kann hart und konfrontativ wirken). Es ist eine eigene Klasse: die strukturelle Lücke zwischen *Form* und *Substanz* iterativer Forschung in einem Single-Agent-Output.

Wir nennen das **Iteratio Simulata**.

Warum es zählt

Multi-Agent-Debate-Frameworks (VirSci, MARG, GAR, MAD) zeigen, wie echte Iteration mit verteilten Rollen aussieht: getrennte Agenten, dokumentierte Eliminationen, externer Richter, Falsifikationsdruck. Single-Agent-Prompts wie der "Autonomous Inventor Mode" imitieren diese Struktur in einem zusammenhängenden Output — und genau das macht sie attraktiv und gefährlich. Sie wirken auf Nicht-Fachleute wie strukturierte Forschung, sind aber linear erzählte Brainstorms mit ausgeschmückter Rahmung.

In der entstehenden LLM-gestützten Forschungspraxis bleibt das Muster systematisch unterdetektiert. Wer einen solchen Output als Eingangsmaterial in eigene Arbeit übernimmt, übernimmt eine Pseudoprozess-Rhetorik mit, ohne sie zu prüfen.

Fünf Marker

M1 — Eliminationsbehauptung ohne Eliminationsspur. Der Output behauptet, Kandidaten verworfen zu haben, benennt sie aber nicht und gibt keine Verwerfungsgründe an.

M2 — Kriterien nach der Bewertung. Bewertungskriterien werden präsentiert, nachdem die Selektion bereits stattgefunden hat, statt vorher fixiert und auf alle Kandidaten gleich angewendet zu werden.

M3 — Reframing etablierter Forschung als Durchbruch. Konzepte, die in der Fachliteratur seit Jahren etabliert sind, werden mit der Rhetorik einer Neuentdeckung präsentiert. Messbar über Abgleich mit Publikationsdaten (Erstpublikation, Reviewdichte) — Methodik aus Afzal et al. (EACL 2026) und der Patent-Novelty-Forschung adaptierbar.

M4 — Selbstvalidierende Rhetorik statt externer Falsifikation. Konfidenz-Adjektive ohne externe Quellen oder Daten, an denen die Bewertung scheitern könnte. Verwandt mit Sycophancy- und Confirmation-Bias-Befunden aus der MAD-Forschung (“Talk Isn’t Always Cheap”, “Peacemaker or Troublemaker”), aber im Single-Agent-Fall.

M5 — Konkretisierungs-Asymmetrie. Über Iterationsschritte nimmt die Konkretheit der Zahlen zu, während die Belegtiefe konstant niedrig bleibt. Präzision wirkt fundiert, ist aber frei gesetzt.

Pre-Registration als Gegenmittel

Die zentrale methodische Frage ist nicht, wie man performative Iteration nachträglich erkennt, sondern wie man sie *vor* der Generierung erschwert. Vier minimale Mechanismen:

- *Pre-Commitment der Kandidatenmenge*: vollständige Kandidatenliste vor Selektion fixieren und hashen
- *Pre-Commitment der Kriterien*: Bewertungskriterien vor Selektion fixieren; spätere Änderungen markieren
- *Externe Verankerung*: mindestens eine indizierte Quelle pro Konzeptklasse (DOI, PubMed-ID, Patentnummer)
- *Asymmetrie-Audit*: Verhältnis quantitativer Aussagen zu belegten quantitativen Aussagen am Ende ausweisen

Diese Mechanismen sind kein vollständiger Schutz, aber sie heben die Schwelle, an der Performativität billig produziert werden kann.

Werkzeug

Die Marker werden in einem eigenständigen Modul *Iterationsprüfer* operationalisiert (Spezifikation v0.4 separat). Nach Phase-0-Verifikation: M1, M2, M5 als strukturelle Heuristiken im Eigenbau; M3 als Adapter zur ICLR-2025-Methodik plus Semantic-Scholar-/OpenAlex-API; M4 als Wrapper um veröffentlichte Sycophancy-Metriken. JSON-Output nach einem werkzeugübergreifenden Analyse-Schema. Anschluss an MYZEL v7 (neues MITRE-Pattern), PRISM v6 (zusätzlicher Frame), RECHERCHE TOOL (Vorfilter).

Was Iteratio Simulata nicht ist

Nicht jeder strukturierte Brainstorm ist *Iteratio Simulata*. Lehrtexte vereinfachen bewusst. Strategiepapiere setzen Prioritäten. Pädagogische Texte reframe etablierte Forschung legitim. Der Unterschied liegt in der **expliziten Markierung als Lehrtext** versus der **Maskerade als Forschungsprozess**. Die Marker sind Hinweise, keine Urteile.

Grenzen

Diese Kurzfassung beruht auf einer Einzelfallstudie (ChatGPT-Brainstorm zu zellfreier Biokatalyse, Mai 2026). Die Marker M1–M5 sind als Hypothesen vorgeschlagen, an weiteren Fällen zu validieren. Modellabhängigkeit ist offen — Vergleichsfälle aus Claude, Gemini, Mistral und offenen Modellen folgen. Pre-Registration-Mechanismus ist skizziert, nicht implementiert.

Selbstanwendung: Die Marker wurden auf die Working Note selbst angewendet. Ergebnis: Die Referenzliste der Vorversion enthielt eine erfundene Autorenschaft, einen falschen Publikationsort und einen weitgehend erfundenen Titel — bei sieben geprüften Einträgen drei Fehler. Der Befund ist in Abschnitt 9 der Working Note offen dokumentiert und stützt Marker M5: Struktur und methodische Sorgfalt in anderen Dimensionen schützen nicht vor unbelegter Präzision.

Position im Forschungsfeld

Das Phänomen liegt im Vorfeld von vier dichten Feldern, deckungsgleich mit keinem:

- **Multi-Agent-Debate-Forschung** liefert den Kontrastmaßstab (echte vs. simulierte Iteration)
- **Sycophancy- und Confirmation-Bias-Forschung in MAD** liefert verwandte Detektionsmetriken für M4
- **Novelty Assessment in Peer Review** liefert die Methodik für M3
- **Selbstkorrektur- und Faithfulness-Forschung** liefert die empirische Grundlage: Huang et al. (ICLR 2024) zeigen, dass Selbstkorrektur ohne externes Feedback strukturell nicht zuverlässig funktioniert

Die beiden nächsten terminologischen Nachbarn sind *unfaithful chain-of-thought* (Turpin et al., NeurIPS 2023) und *potemkin understanding* (Mancoridis et al., ICML 2025). Die Abgrenzung ist präzise: Ersteres betrifft das Verhältnis von Begründung zu Modellinterna, Letzteres das Verhältnis von Konzeptklärung zu Konzeptanwendung. *Iteratio Simulata* betrifft das Verhältnis von behaupteter zu nachweisbarer Prozessstruktur — und ist von diesen drei als einziges allein am Text prüfbar, ohne Modellzugang und ohne Benchmark-Setup. Für die 2026 erschienenen Arbeiten zur performativen Begründungsspur gilt das nicht: Sie messen dieselbe Lücke, brauchen dafür aber Modellzugang und stehen damit neben den Markern, nicht vor ihnen.

Verweise

Vollständige Working Note *Iteratio Simulata v0.4*, Modulspezifikation *Iterationsprüfer v0.4* und Phase-0-Verifikationsbericht: gleicher Autor, gleiches Datum. Lizenz CC BY 4.0. Privat finanziert, keine institutionelle Affiliation.

Thomas Schüller — ORCID 0009-0003-9799-6747 — forschung@tschueller.com — Wien, 2026