

SYNTAGMA FORSCHUNG · FACHBUCHAUSGABE

Iteratio Simulata - Performative Iteration in Sprachmodell-Ausgaben - Syntax-Karte und Lernpfad

Thomas Schüller

ORCID 0009-0003-9799-6747 · FORSCH-PROV0459 · Fachbuchausgabe 1.0 · 14.10.2026

Impressum

Autor	Thomas Schüller, Syntagma Forschung, Wien · ORCID 0009-0003-9799-6747 · forschung@tschueller.com
Kennung	FORSCH-PROV0459
Stand	14.10.2026 · Fachbuchausgabe 1.0
Lizenz	CC BY 4.0
Bezug	syntagma-science.com/werke/iteratio-simulata/

Inhalt

- Anhang S · Syntax-Karte
- Anhang L · Lernpfad
- Glossar

Anhang S · Syntax-Karte

Die Karte gibt das Werk in der Syntagma-Syntax 1.0 wieder (Spezifikation: syntagma-science.com/syntax/). Stufen G0-G5 beschreiben die Beleglage, T1-T6 die Art der Quelle; jede Hypothese nennt ihre Prüfung, jede Karte ihre Grenzen.

Glied	Stufe	Aussage
F1 Frage		Wie lässt sich die Lücke zwischen der Form und der Substanz einer Multi-Agent-Debatte in einem Single-Agent-Output systematisch erkennen, und welche minimalen Vorab-Registrierungen verringern sie vor der Generierung?
F2 Frage		Gibt es im Forschungsumfeld einen Prüfkanal für den einzelnen, nicht instrumentierten Modelllauf, von dem nur der Text bleibt?
G1 Gegenstand		Single-Agent-Outputs großer Sprachmodelle, die verteilte Prüfrollen rhetorisch imitieren ebene: Textoberfläche einzelner Ausgaben, ohne Modellzugang
G2 Gegenstand		Anlassfall: Brainstorm eines Sprachmodells zu zellfreier Biokatalyse über sieben Nutzerturns, eingeleitet durch einen Masterprompt mit mindestens 15 geforderten Iterationen ebene: Einzelfall · zeit: Mai 2026
G3 Gegenstand		Forschungsumfeld des Verfahrensnachweises in LLM-gestützter Wissensarbeit (Feldkarte, vier Erhebungsstränge) ebene: Fachliteratur und Systemarchitekturen · zeit: bis Oktober 2026
B1 Begriff		Iteratio Simulata := strukturelle Lücke zwischen behaupteter und textuell nachweisbarer Prozessstruktur in einem Single-Agent-Output, der eine Multi-Agent-Debatte imitiert fach: llm-evaluation · bruecke: wissenschaftsphilosophie=Cargo-Cult-Science (Form ohne Substanz)
B2 Begriff		Tatsächliche Iteration := nachvollziehbare, nicht-triviale Transformation zwischen zwei Zuständen: Elimination mit Grund, Modifikation mit Angabe, Hinzufügen mit Quelle oder Anwendung eines vorher festgelegten Kriteriums fach: methodologie
B3 Begriff		Unfaithful chain-of-thought := Begründungsspur, die den tatsächlichen Entscheidungsprozess nicht abbildet; Aussage über Modellinterna fach: ml-interpretierbarkeit
B4 Begriff		Potemkin understanding := Kluft zwischen Erklären und Anwenden eines Konzepts bei hoher Benchmark-Leistung fach: llm-evaluation
B5 Begriff		Eliminationsbehauptung ohne Eliminationsspur (M1) := der Output behauptet verworfene Kandidaten, nennt aber weder sie noch Verwerfungsgründe fach: llm-evaluation
B6 Begriff		Kriterien nach der Bewertung (M2) := Bewertungskriterien erscheinen erst, nachdem die Selektion gefallen ist fach: llm-evaluation · bruecke: wissenschaftsmethodik=HARKing (Hypothesizing After the Results are Known)
B7 Begriff		Konkretisierungs-Asymmetrie (M5) := Zahlen, Preise und Zeitlinien werden über die Iterationen konkreter, während die Belegtiefe konstant niedrig bleibt fach: llm-evaluation
B8 Begriff		Vorab-Registrierung := Kandidatenmenge und Kriterien werden vor der Selektion festgelegt und kryptographisch gehasht; spätere Änderungen werden markiert fach: methodologie · bruecke: metawissenschaft=Präregistrierung aus der Replikationskrise
A1 Annahme		Die Marker sind gewichtete Indikatoren auf einem Kontinuum zwischen vollständiger Iteration und vollständiger Performativität, keine binäre Entscheidung
S1 Satz	G3	Große Sprachmodelle korrigieren ihre Argumentation ohne externes Feedback nicht zuverlässig belege: Q1
S2 Satz	G3	Debatten zwischen Agenten können die Genauigkeit über die Zeit senken; Modelle wechseln häufig von korrekten zu inkorrekten Antworten, weil sie Zustimmung über Widerspruch stellen belege: Q2
S3 Satz	G3	Im Anlassfall sind keine 15 Iterationen erkennbar: verworfene Kandidaten werden nicht benannt, die

		Kriterien erscheinen nach der Auswahl, und die Top-Konzepte sind seit zehn bis fünfzehn Jahren in der Literatur etabliert belege: Q6
S4 Satz	G3	Bei der Selbstanwendung waren von achtzehn Referenzangaben fünf fehlerhaft, zwei davon mit erfundener Autorenschaft; alle arXiv-Nummern waren korrekt, fehlerhaft waren nur die Metadaten belege: Q7
S5 Satz	G3	Die erste Werkzeugfassung bewertete einen sorgfältig gearbeiteten Text auf M1 und M2 mit dem Maximalwert; die Fehlerrichtung war dem Zweck entgegengesetzt belege: Q7
S6 Satz	G2	Nach Korrektur erhält der saubere Testtext 0,00 und der Anlassfall 0,80; an zwei Fällen gemessen, ist das kein belastbarer Kennwert, sondern zeigt nur, dass die Marker trennen belege: Q7
S7 Satz	G2	Die Übereinstimmung der inneren Festlegung eines Modells mit der Darstellung in der Begründungsspur liegt bei 61,9 Prozent der Schritte über neun Modelle und sieben Benchmarks belege: Q8
S8 Satz	G3	Iterative Selbstverbesserung kann unter bestimmten Bedingungen messbar funktionieren; nicht jede behauptete Iteration ist simuliert belege: Q9
H1 Hypothese	G2	M5 ist der praktisch wirksamste Marker; er trifft auch dort, wo methodische Sorgfalt in anderen Dimensionen vorhanden ist belege: S4 · pruefung: P1
P1 Prüfung		Fällt, wenn M5 in einem Testkorpus mit unabhängiger menschlicher Vorklassifikation nicht häufiger trifft als die übrigen Marker
H2 Hypothese	G1	Ein leichtgewichtiger Mechanismus (Hash der Eingangsbedingungen, Audit am Ende) genügt, um Performativität deutlich zu verringern pruefung: P2
P2 Prüfung		Fällt, wenn Ausgaben mit Vorab-Registrierung keine deutlich niedrigeren Markerwerte zeigen als vergleichbare Ausgaben ohne
H3 Hypothese	G1	Für eine Vorab-Erklärung des Verfahrens bei KI-gestützter Textarbeit gibt es keinen wissenschaftlichen oder normativen Gebrauch pruefung: P3
P3 Prüfung		Fällt, wenn eine weitere Suche ein solches Instrument für das Manuskript selbst findet, nicht nur auf der Begutachtungsseite
Q1 Quelle	T2	Huang u. a.: Selbstkorrektur von Sprachmodellen ohne externes Feedback, ICLR 2024 nachweis: Huang u. a. 2024
Q2 Quelle	T3	Wynn, Satija und Hadfield: Talk Isn't Always Cheap, Preprint nachweis: Wynn, Satija & Hadfield 2025 · ort: https://doi.org/10.48550/arXiv.2509.05396
Q3 Quelle	T2	Afzal u. a.: Beyond “Not Novel Enough”, EACL 2026; Methodenbasis für M3 nachweis: Afzal u. a. 2026
Q4 Quelle	T2	Turpin u. a.: unfaithful chain-of-thought, NeurIPS 2023 nachweis: Turpin u. a. 2023
Q5 Quelle	T2	Mancoridis u. a.: potemkin understanding, ICML 2025 nachweis: Mancoridis u. a. 2025
Q6 Quelle	T5	Archivierter Wortlaut des Anlassfalls nachweis: Anhang A der Working Note
Q7 Quelle	T5	Selbstanwendung, Referenzprüfung und Gegentests des Werkzeugs nachweis: Abschnitt 9 der Working Note
Q8 Quelle	T3	Li u. a.: When Reasoning Traces Become Performative, Preprint nachweis: Li u. a. 2026 · ort: https://arxiv.org/abs/2605.11746
Q9 Quelle	T2	Madaan u. a.: Self-Refine, NeurIPS 2023 nachweis: Madaan u. a. 2023
M1		Werkzeugbasierte sozio-technische Forschung: Begriffsklärung, Fallanalyse, Operationalisierung,

Methode		Werkzeugentwicklung, Rückkopplung; Abgrenzung gegen vier Nachbardiskurse
M2 Methode		Selbstanwendung der Marker auf die eigene Note; Referenzen an der Primärseite geprüft; Marker an Negativfällen validiert, bevor Positivfälle zählen
L1 Grenze		Begriffsbildende Working Note an einem Einzelfall, keine quantitativ-empirische Studie; der Generalisierungsanspruch ist eingegrenzt
L2 Grenze		Die Abgrenzung zu legitimen rhetorischen Strategien wie Vereinfachung im Lehrtext ist nicht trivial; Nicht-Verbalisierung eines genutzten Hinweises belegt für sich keine Untreue der Spur
L3 Grenze		Die Vorab-Registrierung ist kein vollständiger Schutz; Literatursondierung und Feldkarte sind keine systematische Übersichtsarbeit
O1 Offen		Mindestens drei weitere Fallstudien und ein Testkorpus aus mindestens zwanzig Texten, je zur Hälfte performativ und sorgfältig, aus verschiedenen Modellen und Domänen
O2 Offen		Ob das Phänomen modellabhängig oder modellübergreifend auftritt; Vergleichsfälle aus weiteren Modellen sind nötig
O3 Offen		Die Markerliste M1-M5 ist nicht abgeschlossen, etwa für dialogische Eskalation vom Brainstorm zum Geschäftsplan ohne Validierungsbruch

Verknüpfungen

- S1 *stützt* B1
- S7 *stützt* B1
- S4 *stützt* H1
- S8 *präzisiert* B1
- B8 *überträgt* G1 (L3)
- B5 *verweist auf* `werkzeug:iterationspruefer`
- M1 *verweist auf* `werkzeug:prism`

Anhang L · Lernpfad

Der Lernpfad folgt dem Hausschema von Lern-Agent: Anlass, Aufbau, Abruf, Anwendung, Übertragung, Rückblick. Die Sache steht vor dem Namen. Empfohlene Wiederholung nach 2, 7 und 21 Tagen.

Anlass

1. Vorhersage vor dem Lesen: Wie lässt sich die Lücke zwischen der Form und der Substanz einer Multi-Agent-Debatte in einem Single-Agent-Output systematisch erkennen, und welche minimalen Vorab-Registrierungen verringern sie vor der Generierung?
2. Vorhersage vor dem Lesen: Gibt es im Forschungsumfeld einen Prüfkanal für den einzelnen, nicht instrumentierten Modelllauf, von dem nur der Text bleibt?
3. Gegenstand: Single-Agent-Outputs großer Sprachmodelle, die verteilte Prüfrollen rhetorisch imitieren
4. Gegenstand: Anlassfall: Brainstorm eines Sprachmodells zu zellfreier Biokatalyse über sieben Nutzerturns, eingeleitet durch einen Masterprompt mit mindestens 15 geforderten Iterationen

Aufbau

1. strukturelle Lücke zwischen behaupteter und textuell nachweisbarer Prozessstruktur in einem Single-Agent-Output, der eine Multi-Agent-Debatte imitiert – dafür steht der Begriff „Iteratio Simulata“ (llm-evaluation).
2. nachvollziehbare, nicht-triviale Transformation zwischen zwei Zuständen: Elimination mit Grund, Modifikation mit Angabe, Hinzufügen mit Quelle oder Anwendung eines vorher festgelegten Kriteriums – dafür steht der Begriff „Tatsächliche Iteration“ (methodologie).
3. Begründungsspur, die den tatsächlichen Entscheidungsprozess nicht abbildet; Aussage über Modellinterna – dafür steht der Begriff „Unfaithful chain-of-thought“ (ml-interpretierbarkeit).
4. Kluft zwischen Erklären und Anwenden eines Konzepts bei hoher Benchmark-Leistung – dafür steht der Begriff „Potemkin understanding“ (llm-evaluation).
5. der Output behauptet verworfene Kandidaten, nennt aber weder sie noch Verwerfungsgründe – dafür steht der Begriff „Eliminationsbehauptung ohne Eliminationsspur (M1)“ (llm-evaluation).
6. Bewertungskriterien erscheinen erst, nachdem die Selektion gefallen ist – dafür steht der Begriff „Kriterien nach der Bewertung (M2)“ (llm-evaluation).
7. Zahlen, Preise und Zeitlinien werden über die Iterationen konkreter, während die Belegtiefe konstant niedrig bleibt – dafür steht der Begriff „Konkretisierungs-Asymmetrie (M5)“ (llm-evaluation).
8. Kandidatenmenge und Kriterien werden vor der Selektion festgelegt und kryptographisch gehasht; spätere Änderungen werden markiert – dafür steht der Begriff „Vorab-Registrierung“ (methodologie).

Abruf

1. Ohne Vorlage wiedergeben: Große Sprachmodelle korrigieren ihre Argumentation ohne externes Feedback nicht zuverlässig Wie gut ist das belegt (G3), und womit?
2. Ohne Vorlage wiedergeben: Debatten zwischen Agenten können die Genauigkeit über die Zeit senken; Modelle wechseln häufig von korrekten zu inkorrekten Antworten, weil sie Zustimmung über Widerspruch stellen Wie gut ist das belegt (G3), und womit?
3. Ohne Vorlage wiedergeben: Im Anlassfall sind keine 15 Iterationen erkennbar: verworfene Kandidaten werden nicht benannt, die Kriterien erscheinen nach der Auswahl, und die Top-Konzepte sind seit zehn bis fünfzehn Jahren in der Literatur etabliert Wie gut ist das belegt (G3), und womit?
4. Ohne Vorlage wiedergeben: Bei der Selbstanwendung waren von achtzehn Referenzangaben fünf fehlerhaft, zwei davon mit erfundener Autorenschaft; alle arXiv-Nummern waren korrekt, fehlerhaft waren nur die Metadaten Wie gut ist das belegt (G3), und womit?
5. Ohne Vorlage wiedergeben: Die erste Werkzeuffassung bewertete einen sorgfältig gearbeiteten Text auf M1 und M2 mit dem Maximalwert; die Fehlerrichtung war dem Zweck entgegengesetzt Wie gut ist das belegt (G3), und womit?
6. Ohne Vorlage wiedergeben: Iterative Selbstverbesserung kann unter bestimmten Bedingungen messbar

funktionieren; nicht jede behauptete Iteration ist simuliert Wie gut ist das belegt (G3), und womit?

Anwendung

1. Hypothese: M5 ist der praktisch wirksamste Marker; er trifft auch dort, wo methodische Sorgfalt in anderen Dimensionen vorhanden ist – Welche Beobachtung würde sie zu Fall bringen? (Lösung: Fällt, wenn M5 in einem Testkorpus mit unabhängiger menschlicher Vorklassifikation nicht häufiger trifft als die übrigen Marker)
2. Hypothese: Ein leichtgewichtiger Mechanismus (Hash der Eingangsbedingungen, Audit am Ende) genügt, um Performativität deutlich zu verringern – Welche Beobachtung würde sie zu Fall bringen? (Lösung: Fällt, wenn Ausgaben mit Vorab-Registrierung keine deutlich niedrigeren Markerwerte zeigen als vergleichbare Ausgaben ohne)
3. Hypothese: Für eine Vorab-Erklärung des Verfahrens bei KI-gestützter Textarbeit gibt es keinen wissenschaftlichen oder normativen Gebrauch – Welche Beobachtung würde sie zu Fall bringen? (Lösung: Fällt, wenn eine weitere Suche ein solches Instrument für das Manuskript selbst findet, nicht nur auf der Begutachtungsseite)

Übertragung

1. Im Fach wissenschaftsphilosophie heißt die Sache „Cargo-Cult-Science (Form ohne Substanz)“: Was trägt hinüber, was nicht?
2. Im Fach wissenschaftsmethodik heißt die Sache „HARKing (Hypothesizing After the Results are Known)“: Was trägt hinüber, was nicht?
3. Im Fach metawissenschaft heißt die Sache „Präregistrierung aus der Replikationskrise“: Was trägt hinüber, was nicht?
4. Übertragung B8 → G1: Wo endet sie? (Die Vorab-Registrierung ist kein vollständiger Schutz; Literatursondierung und Feldkarte sind keine systematische Übersichtsarbeit)

Rückblick

1. Wofür gilt das Werk nicht? Begriffsbildende Working Note an einem Einzelfall, keine quantitativ-empirische Studie; der Generalisierungsanspruch ist eingegrenzt
2. Wofür gilt das Werk nicht? Die Abgrenzung zu legitimen rhetorischen Strategien wie Vereinfachung im Lehrtext ist nicht trivial; Nicht-Verbalisierung eines genutzten Hinweises belegt für sich keine Untreue der Spur
3. Wofür gilt das Werk nicht? Die Vorab-Registrierung ist kein vollständiger Schutz; Literatursondierung und Feldkarte sind keine systematische Übersichtsarbeit
4. Offen: Mindestens drei weitere Fallstudien und ein Testkorpus aus mindestens zwanzig Texten, je zur Hälfte performativ und sorgfältig, aus verschiedenen Modellen und Domänen
5. Offen: Ob das Phänomen modellabhängig oder modellübergreifend auftritt; Vergleichsfälle aus weiteren Modellen sind nötig
6. Offen: Die Markerliste M1–M5 ist nicht abgeschlossen, etwa für dialogische Eskalation vom Brainstorm zum Geschäftsplan ohne Validierungsbruch

Glossar

Eliminationsbehauptung ohne Eliminationsspur (M1)	der Output behauptet verworfene Kandidaten, nennt aber weder sie noch Verwerfungsgründe llm-evaluation
Iteratio Simulata	strukturelle Lücke zwischen behaupteter und textuell nachweisbarer Prozessstruktur in einem Single-Agent-Output, der eine Multi-Agent-Debatte imitiert llm-evaluation · ↔ wissenschaftsphilosophie=Cargo-Cult-Science (Form ohne Substanz)
Konkretisierungs-Asymmetrie (M5)	Zahlen, Preise und Zeitlinien werden über die Iterationen konkreter, während die Belegtiefe konstant niedrig bleibt llm-evaluation
Kriterien nach der Bewertung (M2)	Bewertungskriterien erscheinen erst, nachdem die Selektion gefallen ist llm-evaluation · ↔ wissenschaftsmethodik=HARKing (Hypothesizing After the Results are Known)
Potemkin understanding	Kluft zwischen Erklären und Anwenden eines Konzepts bei hoher Benchmark-Leistung llm-evaluation
Tatsächliche Iteration	nachvollziehbare, nicht-triviale Transformation zwischen zwei Zuständen: Elimination mit Grund, Modifikation mit Angabe, Hinzufügen mit Quelle oder Anwendung eines vorher festgelegten Kriteriums methodologie
Unfaithful chain-of-thought	Begründungsspur, die den tatsächlichen Entscheidungsprozess nicht abbildet; Aussage über Modellinterna ml-interpretierbarkeit
Vorab-Registrierung	Kandidatenmenge und Kriterien werden vor der Selektion festgelegt und kryptographisch gehasht; spätere Änderungen werden markiert methodologie · ↔ metawissenschaft=Präregistrierung aus der Replikationskrise