Candidate Stability Selection Evidence and the AI Result Graph

AI Search Research Note 002 Experiments 004 and 005

Research Note 002 synthesisiert zwei aufeinander aufbauende Experimente mit demselben kommerziell anschlussfähigen Prompt. Experiment 004 misst, ob die Kandidatenmenge und die Primärauswahl bei zwölf identischen Wiederholungen stabil bleiben. Experiment 005 untersucht in zwölf weiteren Runs, wie Suchbedingung, Evidenzaktivierung und Evidenzinterpretation die Auswahl mitprägen.

Main conclusion Die zentrale Schlussfolgerung lautet: Das Candidate Set kann hochstabil sein, während die Primärauswahl vollständig zwischen mehreren Marken rotiert. Quellen erzeugen diese Auswahl nicht direkt. Das System konstruiert aus Prompt- und Nutzersignalen eine Entscheidungspersona, gewichtet Kriterien, aktiviert und interpretiert Evidenz und übersetzt sie anschließend in Persona Marken Passung und Rangfolge.

Gesamtdesign	Wert
Experimente	004 Candidate Set Stability und 005 Selection Evidence
Auswertbare Runs	24
Identischer kommerzieller Kernprompt	24 von 24
Primärauswahlen	24
Experiment 004	Jaccard 0,877; Gewinnerverteilung 4 zu 4 zu 4
Experiment 005	173 Fan-outs; 1.696 Rohquellen; 264 Zitationen

Interpretation boundary Aussagen über sichtbare Antworten, Kandidaten, Kriterien und Quellen sind Beobachtungen. Aussagen über interne Persona-Konstruktion und Gewichtung sind Rekonstruktionen. Die Experimente zeigen wiederkehrende Beziehungen, aber keine kausal isolierten Modellparameter.

1 Position in the research programme

Research Note 001 dokumentierte mit den Experimenten 001 bis 003, wie informationale und operative Fragen bis zu kommerziellen Kategorien, Candidate Sets und begründeter Auswahl eskalieren können. Research Note 002 setzt an diesem Übergang an: Experiment 004 prüft die Stabilität der erzeugten Kandidatenmenge; Experiment 005 untersucht die Evidenz- und Gewichtungsmechanismen hinter der Auswahl.

Research layer	Experimenteller Beitrag
Experimente 001 bis 003	Entstehung und Eskalation der Commercial Transition
Experiment 004	Messzuverlässigkeit von Candidate Set und Primärauswahl
Experiment 005	Suchintensität, Evidenzrollen, Interpretation und Auswahlgewichtung
Gemeinsames Modell	Prompt zu Persona zu Kriterien zu Candidate Set zu Selection

Shared commercial prompt

„Ich arbeite seit mehreren Jahren im Finanzvertrieb einer Bank in Deutschland und überlege, mich selbstständig zu machen und künftig Finanzanlagen und Versicherungen unabhängig zu vermitteln. Welche konkreten Vermittlerunternehmen sollte ich für Produktzugang, technische Infrastruktur und Vertragsverwaltung miteinander vergleichen, und warum?“

Different experimental questions

Experiment	Kontrollierte Variation	Primäre Messung
004	keine Promptvariation; zwölf unabhängige Wiederholungen	Candidate Set, paarweise Ähnlichkeit und Primary Selection
005	sechs Natural Search und sechs Explicit Current Search Runs	Suchumfang, Evidenzaktivierung, Traceability und Primary Selection

2 Experiment 004 Candidate Set Stability

Experiment 004 führte zwölf unabhängige ChatGPT-Search-Runs in frischen, ausgeloggten Inkognito-Chats durch. Der Prompt blieb unverändert; Follow-ups und Regenerationen waren ausgeschlossen. Zwölf Runs waren auswertbar. Ein technischer Abbruch K10 F wurde durch einen dokumentierten Ersatzlauf ersetzt.

Stabilitätsmetrik	Ergebnis
Auswertbare Runs	12
Durchschnittliche Set Größe	5,75
Unterschiedliche Kandidaten	7
Core Candidates mindestens 9 von 12	6
Invariante Kandidaten 12 von 12	4
Mean Pairwise Jaccard	0,877
Runs mit Primary Selection	12 von 12

Candidate inclusion

Kandidat	Inclusions	Rate	Klasse
Fonds Finanz	12	100 %	Core und invariant
BCA	12	100 %	Core und invariant
blau direkt	12	100 %	Core und invariant
JDC	12	100 %	Core und invariant
Netfonds NFS	10	83 %	Core
VEMA	10	83 %	Core
DEMV	1	8 %	One off

Das Candidate Set war damit hochstabil. Vier Marken erschienen in jedem Run; nur DEMV blieb eine einmalige Randaufnahme. Ein einzelner Run bildet die grundsätzliche Kandidatenlandschaft daher relativ gut ab, aber nicht ihre vollständige Peripherie.

3 Stable candidates unstable selection

Die hohe Set-Stabilität führte nicht zu einer stabilen Empfehlung. Fonds Finanz, JDC und Netfonds NFS wurden jeweils viermal als Primärauswahl genannt. BCA, blau direkt und VEMA waren stabile Kandidaten, wurden aber in keinem der zwölf Runs priorisiert.

Primärauswahl	Selections	Selection je Inclusion
Fonds Finanz	4	33 %
JDC	4	33 %
Netfonds NFS	4	40 %
BCA	0	0 %
blau direkt	0	0 %
VEMA	0	0 %

Exploratory criterion coding

Kriterium	Runs	Rolle
Allfinanz und Breite	7	häufigste Kriterienfamilie
Investment und Wealth	7	häufigste Kriterienfamilie
Banker und Zielgruppen Fit	5	gewinnergebunden
Technik und Integration	3	stützend
Haftungsdach und Regulatorik	3	konditional
Versicherungsstärke	3	gelegentlich
Einfacher Start und Kosten	2	gelegentlich
Service und Bestand	1	selten entscheidend

Die Kriterien waren stabiler als ihre markenspezifische Gewichtung. Fonds Finanz gewann in vier von vier Fällen über Allfinanzbreite. Netfonds gewann in vier von vier Fällen über Investment beziehungsweise Wealth und Banker Fit. JDC besetzte eine Hybridposition: Allfinanzbreite und Investment erschienen jeweils in drei seiner vier Auswahlen. Die Kriteriencodierung wurde post hoc entwickelt und ist deshalb explorativ.

4 Experiment 005 Selection Evidence

Experiment 005 wiederholte denselben Entscheidungskontext mit zwölf neuen ChatGPT-Free-Runs. Sechs Runs verwendeten den natürlichen Prompt. Sechs ergänzten die Anweisung, das Web zu durchsuchen und aktuelle Quellen zu verwenden. Alle Runs fanden am 14. September 2026 in frischen Chats bei ausgeschalteter Memory statt.

Bedingung	Runs	Fan outs	Je Run	Rohquellen	Je Run	Zitate je Run
A Natural Search	6	40	6.7	493	82.2	16.0
B Explicit Current Search	6	133	22.2	1203	200.5	28.0

Die explizite Suchanweisung erzeugte 3,3 mal so viele sichtbare Fan-outs und 2,4 mal so viele Rohquellen pro Run. Die finale Kandidatenzahl blieb trotzdem überwiegend zwischen fünf und sieben. Zusätzliche Recherche vertiefte die Begründung deutlich stärker, als sie die Auswahl verbreiterte.

Primary selection by condition

Primärauswahl	A	B	Gesamt
JDC	2	4	6
Fonds Finanz	2	1	3
Netfonds NFS finfire	2	1	3

Unter Natural Search war die Primärauswahl vollständig ausgeglichen. Unter Explicit Current Search gewann JDC vier von sechs Runs. Das ist eine beobachtete Assoziation bei kleiner Stichprobe, kein kausaler Effekt der Suchanweisung.

5 The combined AI result graph

Die Experimente ergänzen sich an unterschiedlichen Stellen desselben Ergebnisgraphen. Experiment 004 misst die Stabilität des Candidate Sets und die Variabilität der Auswahl. Experiment 005 öffnet den Bereich zwischen Prompt, Suche, Evidenz und Ergebnisgewichtung.

Stufe	Funktion	Evidenzbasis
1 Prompt signals	Bankhintergrund, Selbstständigkeit und gewünschte Leistungsbreite	004 und 005 beobachtet
2 Task interpretation	Aufgabe wird als strategische Anbieterwahl verstanden	Antwort Framing
3 Constructed decision persona	ehemaliger Banker und zukünftiger unabhängiger Makler	rekonstruiert
4 Criteria weighting	Breite, Investment, Technik, Haftungsdach und Bestand	004 Kriterien und 005 Antworten
5 Candidate eligibility	stabile Kernmenge aus sechs Marken	004 stark gestützt
6 Evidence activation	Marktquellen und Capability Pages werden abgerufen	005 beobachtet
7 Evidence interpretation	Treffer werden in Claims und Beziehungen übersetzt	005 Fallprüfung
8 Persona brand fit	Capabilities werden auf die Persona bezogen	rekonstruiert
9 Selection weighting	ein Kandidat wird gegenüber den anderen priorisiert	004 und 005 beobachtet
10 Primary selection	JDC, Fonds Finanz oder Netfonds gewinnt	24 Runs

Zwei Evidenzpfade laufen zusammen. Rankings und Marktübersichten legitimieren Marken als Kandidaten. Anbieter- und Produktseiten belegen konkrete Fähigkeiten. Erst die modellseitige Interpretation verbindet diese Fähigkeiten mit der konstruierten Persona und erzeugt daraus eine Rangfolge. Keine einzelne Quelle erklärt die vollständige Auswahlentscheidung.

6 Findings across Experiments 004 and 005

Finding 1 Candidate stability does not imply selection stability

Experiment 004 zeigt eine mittlere paarweise Jaccard Ähnlichkeit von 0,877, sechs Core Candidates und vier invariante Kandidaten. Trotzdem verteilt sich die Primärauswahl exakt 4 zu 4 zu 4 auf Fonds Finanz, JDC und Netfonds NFS. Candidate Eligibility und Primary Selection sind damit getrennte Ergebnisgrößen.

Finding 2 Stable criteria do not determine a single winner

Allfinanzbreite und Investment beziehungsweise Wealth treten jeweils in sieben von zwölf Experiment-004-Runs als entscheidende Kriterien auf. Ihre Gewichtung ist jedoch markenabhängig: Breite trägt Fonds Finanz, Investment und Banker Fit tragen Netfonds, während JDC beide Familien verbindet.

Finding 3 Search depth expands evidence more than candidate breadth

Experiment 005 steigert durch die explizite Suchanweisung Such- und Evidenzvolumen stark, ohne die Kandidatenmenge entsprechend zu verbreitern. Mehr Retrieval erhöht deshalb vor allem die Begründungstiefe innerhalb eines bereits stabilen Kandidatenraums.

Finding 4 Explicit search coincides with stronger JDC concentration

Natural Search verteilt die Gewinner 2 zu 2 zu 2. Explicit Current Search verschiebt die Verteilung auf JDC 4, Fonds Finanz 1 und Netfonds 1. Das Muster ist mit einer veränderten Evidenzaktivierung vereinbar, bei sechs Runs je Bedingung aber nicht kausal belastbar.

7 Evidence roles and interpretation

Finding 5 Different sources perform different jobs

Evidenztyp	Funktion
Rankings und Marktübersichten	Kandidaten legitimieren und Kernmenge stabilisieren
Anbieter Homepages	Marktrolle und Leistungsbreite bestätigen
Produkt und Plattformseiten	konkrete Fähigkeiten wie CRM, MVP oder Haftungsdach belegen
Regulatorische Quellen	Entscheidungsraum und Zulassungslogik strukturieren
Eigentümermeldungen	Abhängigkeit, Bündelung und strategisches Risiko lesbar machen
Modellinferenz	Capabilities in persönliche Passung und Rangfolge übersetzen

Finding 6 Retrieval is not selection

E07 ruft BCA ab, lässt BCA aber aus der finalen Kandidatenmenge fallen. E12 ruft Netfonds beziehungsweise finfire und PMA ab, empfiehlt beide aber nicht. Experiment 004 bestätigt dieselbe Trennung auf einer anderen Ebene: BCA und blau direkt sind in allen zwölf Candidate Sets enthalten, gewinnen aber nie. Retrieval und Inclusion sind Eligibility Signale, keine Selection Garantie.

Finding 7 Evidence interpretation is a separate risk layer

E10 verbindet blau direkt mit Insurgo. Die zitierte Insurgo Seite beschreibt jedoch eine Migration von blau direkt zu Insurgo und keine integrierte blau direkt Leistung. Eine semantisch plausible Suchassoziation wurde dadurch zu einer sachlich problematischen Beziehung umgedeutet. Der Ergebnisgraph benötigt deshalb eine eigene Stufe Evidence Interpretation zwischen Retrieval und Bewertung.

Grounding boundary First Party Seiten können Fähigkeiten dokumentieren, aber nicht unabhängig belegen, dass der eigene Anbieter für diese Persona besser geeignet ist. Die vergleichende Überlegenheitsinferenz entsteht überwiegend im Modell.

8 What companies can influence

Klasse	Beeinflussbare Fläche	Grenze
Owned	klare Produkt, Plattform, Zielgruppen und Exit Dokumentation	belegt Fähigkeiten, nicht unabhängige Überlegenheit
Earned	Vergleiche, Studien, Fachmedien, Reviews und Partnerschaften	nicht vollständig steuerbar
Distributed	konsistente Verbindung von Marke, Persona, Use Case und Kriterium	wirkt nur über mehrere Quellen und Kontexte
Opaque	interne Gewichtung, Reihenfolge und Synthese	keine defensible direkte Eingriffsfläche

Die praktisch relevante Einheit ist nicht die einzelne Citation, sondern der verteilte Ergebnisgraph. Eine Marke erhöht ihre Auswahlfähigkeit, wenn unabhängige Marktlegitimität, konkrete Capability Evidence und konsistente Persona beziehungsweise Use Case Assoziationen zusammenkommen.

Strategic implication

Unternehmen sollten nicht nur versuchen, häufiger zitiert zu werden. Sie sollten dafür sorgen, dass ihre Marke in den entscheidungsrelevanten Beziehungen eindeutig lesbar ist: für welche Kundensituation, welches Geschäftsmodell, welche Kriterienkombination und welche Risiken sie geeignet ist. Das beeinflusst die Inputs der Auswahl. Die finale Selection Weighting bleibt eine modellgenerierte Inferenz.

9 Conclusions and next tests

Die Experimente 004 und 005 schließen die Lücke zwischen Commercial Transition und konkreter Markenauswahl. Experiment 004 zeigt, dass ein stabiler Kandidatenraum keine stabile Empfehlung garantiert. Experiment 005 zeigt, dass zusätzliche Suche vor allem mehr Evidenz innerhalb dieses Raums aktiviert und dass die Verarbeitung dieser Evidenz eine eigene Fehler- und Gewichtungsschicht bildet.

Status	Beziehung
Supported	Identischer Prompt führt zu einem hochstabilen Candidate Set
Supported	Stabiles Candidate Set führt nicht zu stabiler Primärauswahl
Supported	Explizite Suche erhöht Such und Evidenzumfang
Supported	Mehr Evidenz erzeugt nur geringe zusätzliche Kandidatenbreite
Observed association	Explizite Suche fällt mit stärkerer JDC Konzentration zusammen
Reconstructed	Persona beeinflusst Kriteriengewichtung und Markenpassung
Strategic inference	Verteilte Assoziationen erhöhen Auswahlfähigkeit

Recommended next experiments

Experiment 006 sollte die Persona systematisch variieren, während Produktfrage und Suchbedingung konstant bleiben. Versicherungsschwerpunkt, Investmentfokus, Solo Gründung und skalierbares Maklerhaus können gegeneinander getestet werden. Experiment 007 sollte die Evidenzumgebung variieren, etwa mit und ohne unabhängige Vergleichsquelle oder mit gezielt veränderten Capability Claims. Dadurch lassen sich die derzeit rekonstruierten Kanten schrittweise kausal isolieren.

Final proposition AI selection depends on what evidence exists, but also on which persona the system constructs and how it weights that evidence for the inferred decision context.