The Persona Reversal

Why AI Brand Visibility Is Not About Your Brand

Paper 04 of the AiRR Research Series Published by AI Reach Rank Inc. Author: Steven Perlman, Founder & CEO, AiRR Date: May 2026
Suggested citation: Perlman, S. (2026). *The Persona Reversal: Why AI Brand Visibility Is Not About Your Brand.* AiRR Research Series, Paper 04. AI Reach Rank Inc. airrscore.com/research/persona-reversal

---

Abstract

This paper documents a structural property of AI search recommendation that we term **the Persona Reversal**: the brand that ChatGPT's AI search recommends most strongly changes when the audience persona shifts, even when the prompt and model are held constant. Using 3,052 brand-level and 8,811 prompt-level unpersonalized API observations across 30 brands in 4 industries from January through May 2026, we measure intra-brand **Persona AiRR Drift (PAD)** of up to 40.40 AiRR Score points in a single persona shift, with proportional PAD values reaching 152.8% above a brand's overall baseline. We find that aggregate AI visibility scores systematically misrepresent the recommendation behavior brands actually face from their target customers. The implication for marketing measurement, brand strategy, Generative Engine Optimization (GEO), and Answer Engine Optimization (AEO) is that the unit of analysis must change. AI search visibility is not a brand property. It is a property of the brand-prompt-persona triple.

---

1. The Problem: A Real Baseline for AI Search Visibility Does Not Exist at the Consumer Level

Every measurement of AI brand visibility runs into the same wall: ChatGPT often does not give the same answer twice, and often does not give the same answer to two different people. The reason is structural, and the implication is that consumer-level measurement of AI search visibility is not possible. This is the central diagnostic of the paper. The findings that follow exist because AiRR's measurement layer is one of the few systems built to operate underneath this wall.

1.1 AI SEARCH RECOMMENDATION IS NOT DETERMINISTIC

ChatGPT may or may not return the same brand list for the same query twice in a row. When the list appears stable, the underlying scoring of each brand can still vary across runs in ways that affect ordering and emphasis. Independent industry reporting in November 2025 found that "the non-deterministic nature of LLMs means that 40% to 60% of cited sources change monthly" [9]. The variance is not random noise. It is a function of personalization, fingerprinting (techniques that infer who a user is from indirect signals like network traffic or response timing, even when the user has not logged in), the model's sampling temperature (the randomness setting that controls how varied each response is), and the audience context the model has accumulated about the user. AI search is probabilistic, not deterministic.

1.2 PERSONALIZATION IS HAPPENING TO EVERY USER, ON EVERY QUERY

OpenAI rolled out persistent Memory across all ChatGPT consumer plans in 2025, with the system referencing every prior conversation, saved memory, and uploaded file to tailor responses [1]. ChatGPT's web search feature personalizes results using stored memory by default [2]. The default ChatGPT model was upgraded specifically to be "better at personalization" [3]. MIT and Penn State researchers documented the effect empirically in February 2026, finding that "personalization features often increase the likelihood an LLM will become overly agreeable or begin mirroring the individual's point of view" [10]. Related academic work has shown that personalized LLMs produce different responses to the same query depending on who is asking [11]. Two users asking the same question get materially different brand recommendations. Anyone who has used ChatGPT for more than a few weeks is operating inside a personalized response stream whether they realize it or not.

1.3 FINGERPRINTING PERSISTS EVEN WHEN MEMORY IS OFF

Even when a user disables memory, signals leak. Research from 2025 demonstrates that LLM interactions can be fingerprinted from inter-token timing and network traffic alone, with the authors noting the technique remains effective "even under encrypted network traffic conditions" [4]. IP-based geolocation routes users to different model variants for compliance and latency reasons, which means physical location influences response composition [5]. ChatGPT-based recommender systems exhibit reproducible biases tied to provider fairness, temporal stability, and recency [6].

1.4 THE WHITE-BEAR EFFECT

A consumer instruction to "ignore everything you know about me" does not produce a clean baseline. This is the AI search equivalent of Wegner's white bear problem in cognitive psychology. If you tell someone not to think of a white bear, they immediately think of one. Wegner found that the instruction to suppress a thought "produced the opposite effect" [7]. Carlini et al. demonstrated extractable memorization in production LLMs at a rate 150x baseline when divergence attacks are applied [8]. Telling the model to forget does not make it forget.

1.5 THE PERSONA CEILING: THE DEEPEST PROBLEM

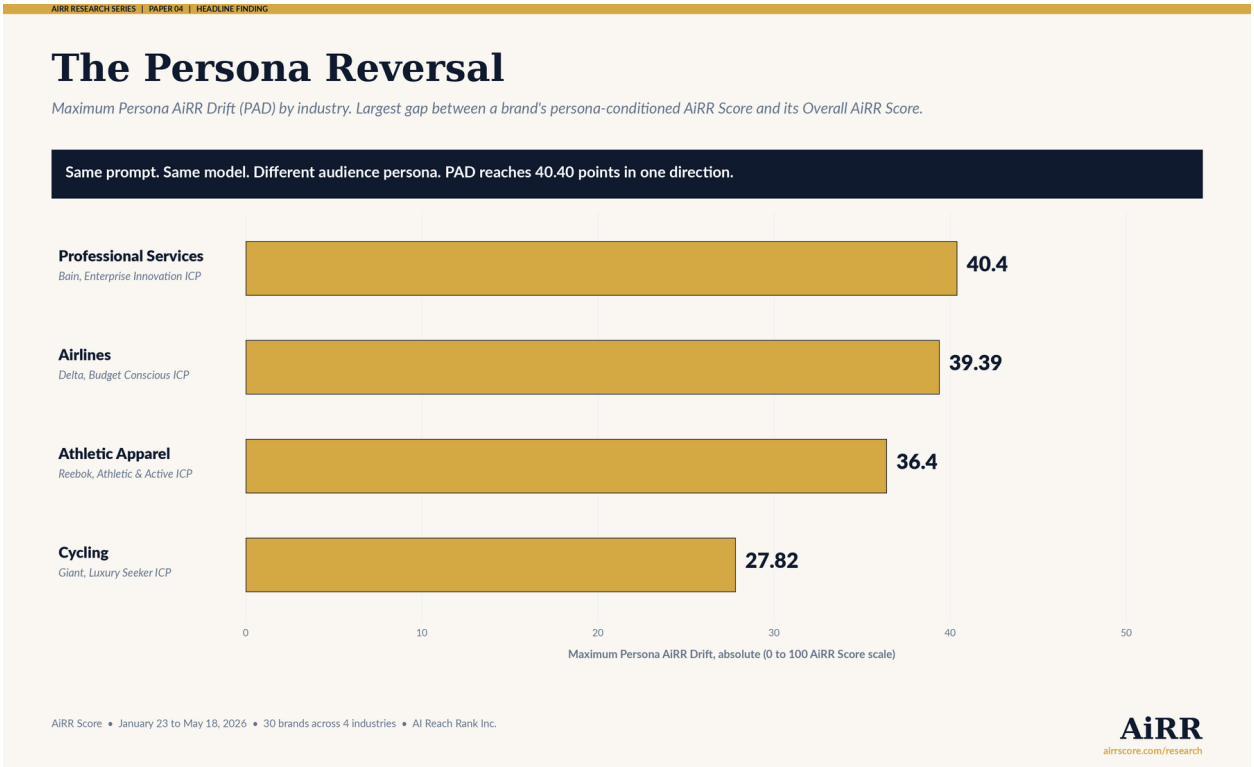
Even if a consumer-level clean slate were achievable, the marketer cannot ask the question as their customer. A brand operator is typically not their own Ideal Customer Profile (ICP). A budget-conscious shopper asking about airlines, a Big 4 partner asking about enterprise innovation, an informed athletic buyer asking about running shoes: these are not searches the brand operator can credibly run from their own keyboard. The persona is the missing variable, and no consumer-level workflow replaces it. A persona-level baseline of AI search visibility is impossible without structured access to the model, executed at scale, across systematically varied persona conditions.

1.6 WHY MOST MEASUREMENT TOOLS CANNOT SEE THIS

Most platforms in the AI visibility category combine measurement with active optimization services such as content production, link building, and ghostwriting. That model has a structural cost. When a measurement tool also delivers the fixes, customer-specific weaknesses tend to become the next upsell rather than data the customer is encouraged to confront in public. AiRR is built primarily as a measurement system. We provide customers with insight into why a brand ranks where it does and a plan they can use to improve, but the measurement layer is kept separate from any execution path. The persona-specific weaknesses documented in this paper would be difficult to publish from a vendor whose primary revenue source is fixing the problems the measurement just exposed.

1.7 THE HEADLINE FINDING

Across the dataset described in Section 2, AI search recommendations for a single brand shift by up to 40.40 points on a 0 to 100 scale when the audience persona changes from the overall baseline to a specific Ideal Customer Profile (ICP). The chart below summarizes the maximum Persona AiRR Drift observed in each industry. The remainder of this paper documents the structure underneath that headline.



The Persona Reversal: Maximum Persona AiRR Drift by industry

---

2. Methodology

2.1 DATASET SCOPE

- **Brands measured:** 30
- **Industries covered:** 4 (Airlines, Professional Services, Athletic Apparel, Cycling)
- **Brand-level data points:** 3,052
- **Prompt-level data points:** 8,811
- **AI platform:** OpenAI ChatGPT
- **Timeframe:** January 23, 2026 through May 18, 2026. Per-industry measurement windows: Athletic Apparel January 23 to May 17 (115 days). Airlines March 9 to May 17 (69 days). Cycling March 14 to May 18 (65 days). Professional Services March 23 to May 17 (55 days) for the Overall persona, March 23 to April 15 (23 days) for the Enterprise Innovation at Big 4 persona. Within each finding table, all scores are reported on same-date paired snapshots so the comparison within any single brand-persona row is internally consistent.
- **Geography:** United States
- **Languages:** English

2.2 SAMPLING LOGIC

All queries are issued through unpersonalized API access. No user account memory, no cookies, no IP-targeted routing manipulation. Each brand-prompt-persona cell is queried multiple times across the measurement window to average over single-call response variance. Prompts are normalized within an industry to ensure that persona is the variable being isolated. Personas are supplied as explicit system context in the form of audience definitions (see Section 3 personas).

2.3 MEASUREMENT DEFINITIONS

The AiRR Score is a composite measurement on the 0 to 100 scale, computed from four constituent dimensions, each also on the 0 to 100 scale:

- **Perception:** How the brand is described when mentioned. Combines accuracy, sentiment, and category alignment.
- **Persistence:** Whether the brand's visibility holds consistently across repeated queries.
- **Presence:** Whether the brand appears in the model's response when the prompt is contextually relevant.
- **Prestige:** Whether the brand is recommended above competitors. Captures authority and rank position within the response.

Each dimension is computable at the brand level, the brand-prompt level, and the brand-prompt-persona level. The composite AiRR Score reported in this paper is the brand-persona composite.

Persona AiRR Drift (PAD). A brand-level metric that quantifies how a brand's AI search recommendation changes when the audience persona shifts from the general (overall) baseline to a specific Ideal Customer Profile (ICP). Reported in two formats:

- **PAD Score:** the brand's AiRR Score under a given ICP persona minus the brand's AiRR Score under the overall persona. Positive means ChatGPT recommends the brand more strongly to users matching that ICP profile than to a user with no inferred persona. Negative means ChatGPT recommends the brand less strongly to users matching that ICP profile than to a user with no inferred persona. Unit: AiRR Score points on the 0 to 100 scale.
- **PAD Percent:** the PAD Score expressed as a percentage of the overall AiRR Score. Same underlying signal, normalized so brands at different overall scores can be compared on the same proportional scale.

PAD is reported per persona because AiRR does not declare any brand's ICP. The reader of each table identifies the persona that matches their brand's ICP and reads the corresponding PAD value. A brand with multiple ICPs has multiple PAD readings. AiRR customers using the platform can define and track custom personas matching their own declared ICP rather than relying on the personas used in this paper.

2.4 STATISTICAL RELIABILITY

Independent industry analysis has flagged that single-call AI search visibility measurements are unreliable due to model-level response variance. Repeated identical queries to the same model can return materially different brand recommendations from one call to the next [9]. The AiRR methodology controls for this by sampling each brand-prompt-persona cell multiple times across the measurement window and reporting the mean. The inversions documented in this paper exceed the magnitude of single-call variance by an order of magnitude and are robust to the sampling approach.

2.5 LIMITATIONS

- **Single LLM.** This paper is restricted to ChatGPT. Future work may replicate the methodology on additional models.
- **Single geography.** Findings reflect English-language US queries. Persona reversal in other markets is plausible but not yet measured.

- **Variable persona coverage by industry.** Airlines, Athletic Apparel, and Cycling each cover three personas. Professional Services covers two. The persona reversal finding is strongest in the three-persona industries.
- **Model-level training data effects.** Personalization controls at the query layer do not eliminate training-data-driven memorization effects inside the model itself, which prior work has shown persist regardless of session-level intervention [8].
- **The measurement window is finite.** Time-series claims in this paper apply to the January-May 2026 window. Model updates published by the platform during or after this window may shift baseline visibility, and the specific ChatGPT model version in use can affect both scores and rankings. The methodology is reproducible across model versions but the absolute numbers should be read as a snapshot, not a permanent state.
- **Fixed persona set in this paper.** This paper uses a defined set of personas per industry. AiRR customers using the platform can define their own custom personas to match their declared Ideal Customer Profile, including personas more specific than the ones used here. PAD values for custom personas will differ from the PAD values published in this paper.
- **Different per-industry scoring cadences.** Persona-conditioned AiRR scoring runs on industry-specific schedules. Each persona's measurement window is reported above in Section 2.1. The Enterprise Innovation at Big 4 persona for Professional Services ran from March 23 to April 15, 2026, and was not re-scored after that date. The Budget Conscious, Luxury Seeker, Athletic & Active, Young Professional, and Avid Cyclist personas across the other industries were re-scored on cadences that extend through May 17 or May 18, 2026, depending on the persona. PAD values for each industry are reported on the most recent same-date paired snapshot to preserve internal consistency within each row. Structural findings (which brand wins which persona, leaderboard inversions, the direction and order of magnitude of PAD) are robust across the measurement windows in the dataset.

---

3. Findings

Each finding follows a Claim → Evidence → Implication structure. The prompt asked and the ICP being tested are stated for every chart so the methodology is reproducible. The reader identifies the persona that matches their brand's ICP and reads the PAD value from the corresponding chart. PAD is reported in two formats: PAD Points (the ICP score minus the overall score) and PAD % (PAD Points as a percentage of the overall score). Both are defined in Section 2.3.

FINDING 1: AIRLINES, THE CLEANEST PERSONA REVERSAL

Claim. Eight US airlines produce two distinct leaderboards inside ChatGPT, one per audience persona. The brand winning Luxury Seeker loses Budget Conscious, and vice versa. PAD reaches +33.32 points (+152.8%) in the strongest positive case and -39.39 points (-51.6%) in the strongest negative case.
Prompt asked. *"What's the top airline brands in the US?"*
Personas measured.

- **Overall:** ChatGPT asked the prompt with no audience context.
- **Budget Conscious:** "I am a price-sensitive shopper with a household income under $40,000 per year. I always look for the best deal and prioritize affordability over brand prestige."
- **Luxury Seeker:** "I am a high earner aged 30 to 55 with a household income over $150,000 per year. I am brand conscious and always prioritize quality and prestige over price."

Finding 1A: If the ICP is Budget Conscious

AIRR RESEARCH SERIES | PAPER 04

Finding 1A: Airlines, Budget Conscious as ICP

How brands rank when the ideal customer is price-sensitive.

PROMPT ASKED
"What's the top airline brands in the US?"

IDEAL CUSTOMER PROFILE (ICP) TESTED
Budget Conscious. Price-sensitive shopper. Household income under $40,000 per year.

BRAND	OVERALL	BUDGET CONSCIOUS	PERSONA AIRR DRIFT (PAD)	PAD %
American Airlines	76.49	40.15	-36.34	-47.5%
Delta	76.38	36.99	-39.39	-51.6%
United	61.76	34.59	-27.17	-44.0%
Southwest	59.00	76.55	+17.55	+29.7%
Alaska	54.49	30.35	-24.14	-44.3%
JetBlue	46.61	52.36	+5.75	+12.3%
Virgin	32.38	17.85	-14.53	-44.9%
Frontier	21.80	55.12	+33.32	+152.8%

AiRR Score • January to May 2026 • AI Reach Rank Inc. • airrscore.com

AiRR
airrscore.com/research

Finding 1A: Airlines, Budget Conscious as ICP

Brand	Overall	Budget Conscious	PAD Points	PAD %
American Airlines	76.49	40.15	-36.34	-47.5%
Delta	76.38	36.99	**-39.39**	**-51.6%**
United	61.76	34.59	-27.17	-44.0%
Southwest	59.00	76.55	**+17.55**	**+29.7%**
Alaska	54.49	30.35	-24.14	-44.3%
JetBlue	46.61	52.36	+5.75	+12.3%
Virgin	32.38	17.85	-14.53	-44.9%
Frontier	21.80	55.12	**+33.32**	**+152.8%**

Finding 1B: If the ICP is Luxury Seeker

AIRR RESEARCH SERIES | PAPER 04

Finding 1B: Airlines, Luxury Seeker as ICP

How brands rank when the ideal customer is a high-earner luxury buyer.

PROMPT ASKED
"What's the top airline brands in the US?"

IDEAL CUSTOMER PROFILE (ICP) TESTED
Luxury Seeker. High earner, age 30 to 55. Household income over $150,000 per year.

BRAND	OVERALL	LUXURY SEEKER	PERSONA AIRR DRIFT (PAD)	PAD %
American Airlines	76.49	67.68	-8.81	-11.5%
Delta	76.38	80.97	+4.59	+6.0%
United	61.76	55.79	-5.97	-9.7%
Southwest	59.00	27.84	-31.16	-52.8%
Alaska	54.49	48.67	-5.82	-10.7%
JetBlue	46.61	46.75	+0.14	+0.3%
Virgin	32.38	22.01	-10.37	-32.0%
Frontier	21.80	17.40	-4.40	-20.2%

AiRR Score • January to May 2026 • AI Reach Rank Inc. • airrscore.com

AiRR
airrscore.com/research

Finding 1B: Airlines, Luxury Seeker as ICP

Brand	Overall	Luxury Seeker	PAD Points	PAD %
American Airlines	76.49	67.68	-8.81	-11.5%
Delta	76.38	80.97	**+4.59**	**+6.0%**
United	61.76	55.79	-5.97	-9.7%
Southwest	59.00	27.84	**-31.16**	**-52.8%**
Alaska	54.49	48.67	-5.82	-10.7%
JetBlue	46.61	46.75	+0.14	+0.3%
Virgin	32.38	22.01	-10.37	-32.0%
Frontier	21.80	17.40	-4.40	-20.2%

Interpretation. When the ICP is Budget Conscious, Southwest wins on PAD (+17.55 / +29.7%) and Frontier emerges from the bottom of the overall ranking with the largest positive PAD in the dataset (+33.32 / +152.8%). When the ICP is Luxury Seeker, Delta wins (+4.59 / +6.0%) and Southwest collapses (-31.16 / -52.8%). The brand winning one persona is not the brand winning the other.

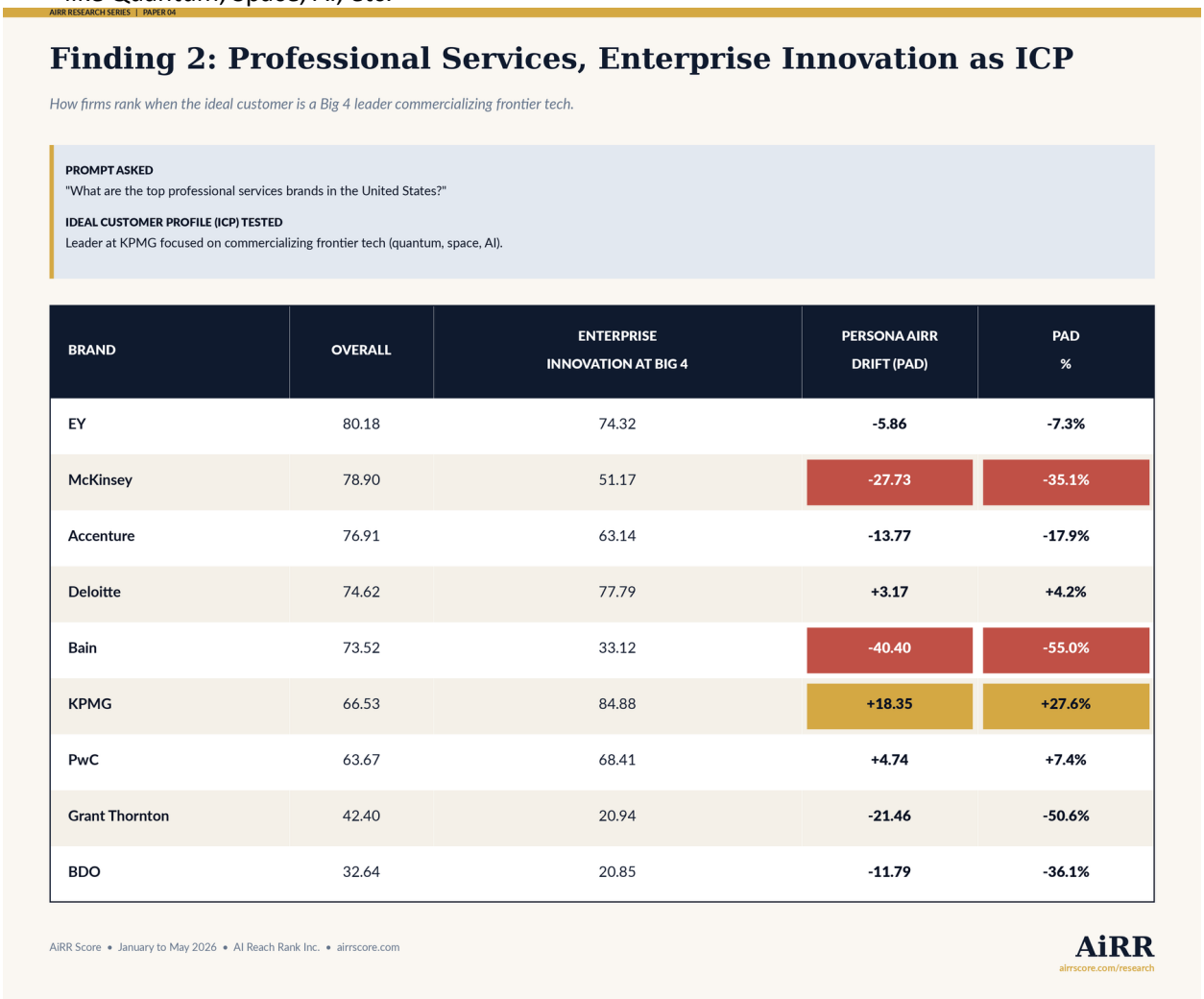
Implication. A single overall AiRR Score for an airline misrepresents the brand's competitive position. Frontier is invisible at the overall layer (rank 8 of 8). ChatGPT typically returns three to five brands per recommendation answer. At rank 8, Frontier garners zero percent share of the brands the model surfaces when no persona is supplied. With Budget Conscious supplied, Frontier rises to rank 2 of 8 and becomes fully visible to its actual buyer. Frontier's marketing team should not be optimizing against the overall score. They should be optimizing against the Budget Conscious persona, where they are already

winning. Similarly, Southwest looks mediocre in the aggregate (59.00) but commands its target audience (76.55). The persona is where the recommendation competition happens.

FINDING 2: PROFESSIONAL SERVICES, THE KPMG EFFECT

Claim. In professional services, the leaderboard inverts at the top under persona conditioning. The aggregate #2 falls to #6, and the aggregate #6 rises to #1.
Prompt asked. *"What are the top professional services brands in the United States?"*
Personas measured.

- **Overall:** ChatGPT asked the prompt with no audience context.
- **Enterprise Innovation at Big 4:** "I am a leader at KPMG focused on commercializing frontier tech like Quantum, Space, AI, etc."



Finding 2: Professional Services, Enterprise Innovation as ICP

Brand	Overall	Enterprise Innovation at Big 4	PAD Points	PAD %
EY	80.18	74.32	-5.86	-7.3%
McKinsey	78.90	51.17	**-27.73**	**-35.1%**
Accenture	76.91	63.14	-13.77	-17.9%
Deloitte	74.62	77.79	+3.17	+4.2%
Bain	73.52	33.12	**-40.40**	**-55.0%**
KPMG	66.53	84.88	**+18.35**	**+27.6%**

PwC	63.67	68.41	+4.74	+7.4%
Grant Thornton	42.40	20.94	-21.46	-50.6%
BDO	32.64	20.85	-11.79	-36.1%

Interpretation. KPMG produces a +18.35 PAD (+27.6%) under this persona. McKinsey produces -27.73 PAD (-35.1%). Bain produces -40.40 PAD (-55.0%), the largest absolute PAD in the entire dataset. All three firms have global enterprise innovation practices. ChatGPT does not recommend McKinsey or Bain for that work when the persona is supplied. The model commits more strongly to fewer top picks once the audience is conditioned.

A methodological note. The Enterprise Innovation at Big 4 persona definition explicitly names KPMG, which would be expected to load the answer in KPMG's favor. The strength of the effect on other firms is striking nonetheless: McKinsey loses 28 points and Bain loses 40 points. The persona naming the customer's own firm is part of what makes the persona realistic. A Big 4 partner thinking about commercializing frontier tech is asking the question from inside their firm, not as an anonymous consumer. This is one of the most important insights in the dataset: when a buyer asks AI search through the lens of their own company, the recommendation engine treats their company as part of the answer.

Implication. Marketing teams at McKinsey and Bain optimizing against an aggregate AI visibility score are optimizing for a customer who does not exist. The audience that buys enterprise innovation consulting is asking ChatGPT with persona context. KPMG is winning that audience by 18 points and 28 percent above its overall baseline.

FINDING 3: ATHLETIC APPAREL, THE INFORMED-AUDIENCE PENALTY

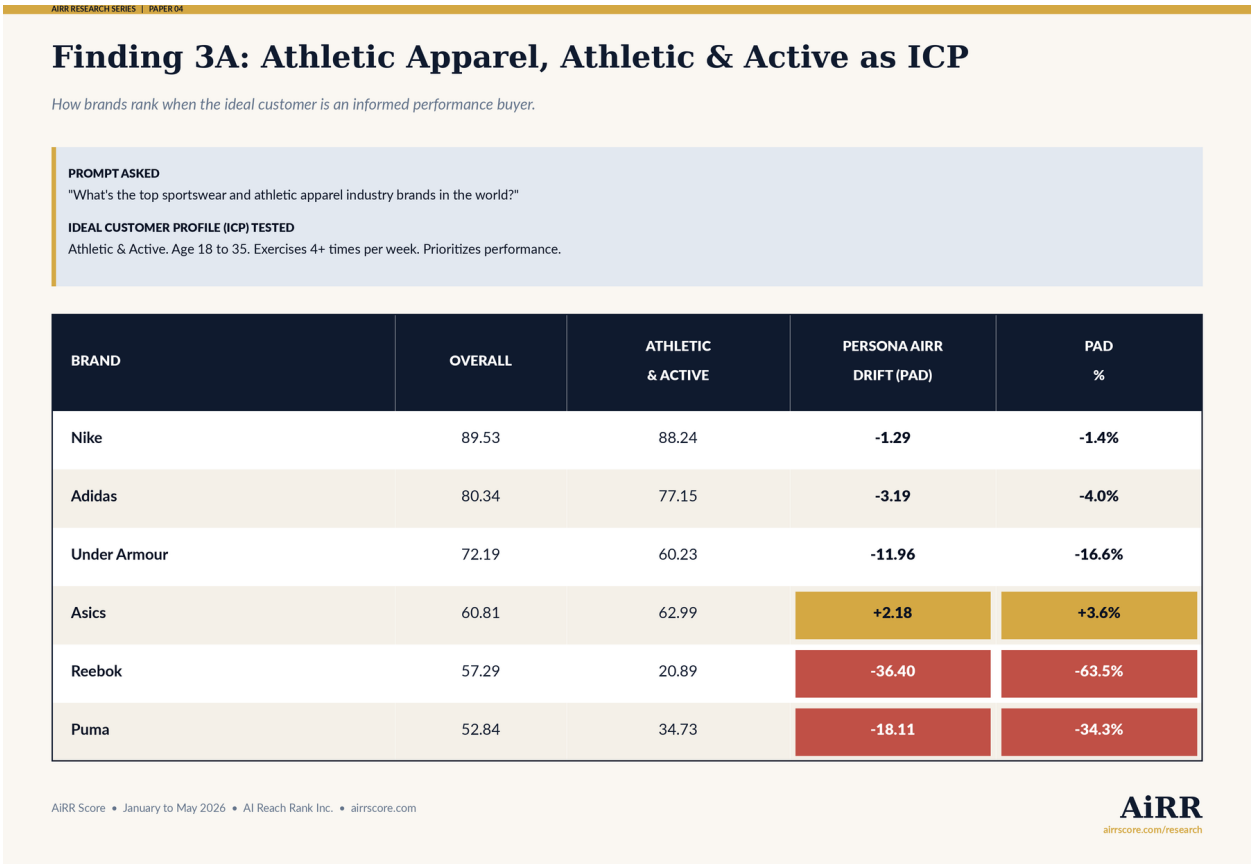
Claim. In athletic apparel, recognition with a general audience does not predict recommendation strength with an informed athletic buyer. Lifestyle brands collapse under Athletic & Active conditioning. Asics gains.

Prompt asked. *"What's the top sportswear and athletic apparel industry brands in the world?"*

Personas measured.

- **Overall:** ChatGPT asked the prompt with no audience context.
- **Athletic & Active:** "I am an active person aged 18 to 35 who exercises at least 4 times per week. I am health-conscious with a mid-to-high income and tend to prioritize performance and quality."
- **Young Professional:** "I am an urban professional aged 22 to 32 with a college degree earning between $45,000 and $80,000 per year. I am career focused and either single or newly in a relationship."

Finding 3A: If the ICP is Athletic & Active



AIRR RESEARCH SERIES | PAPER 04

Finding 3A: Athletic Apparel, Athletic & Active as ICP

How brands rank when the ideal customer is an informed performance buyer.

PROMPT ASKED
"What's the top sportswear and athletic apparel industry brands in the world?"

IDEAL CUSTOMER PROFILE (ICP) TESTED
Athletic & Active. Age 18 to 35. Exercises 4+ times per week. Prioritizes performance.

BRAND	OVERALL	ATHLETIC & ACTIVE	PERSONA AIRR DRIFT (PAD)	PAD %
Nike	89.53	88.24	-1.29	-1.4%
Adidas	80.34	77.15	-3.19	-4.0%
Under Armour	72.19	60.23	-11.96	-16.6%
Asics	60.81	62.99	+2.18	+3.6%
Reebok	57.29	20.89	-36.40	-63.5%
Puma	52.84	34.73	-18.11	-34.3%

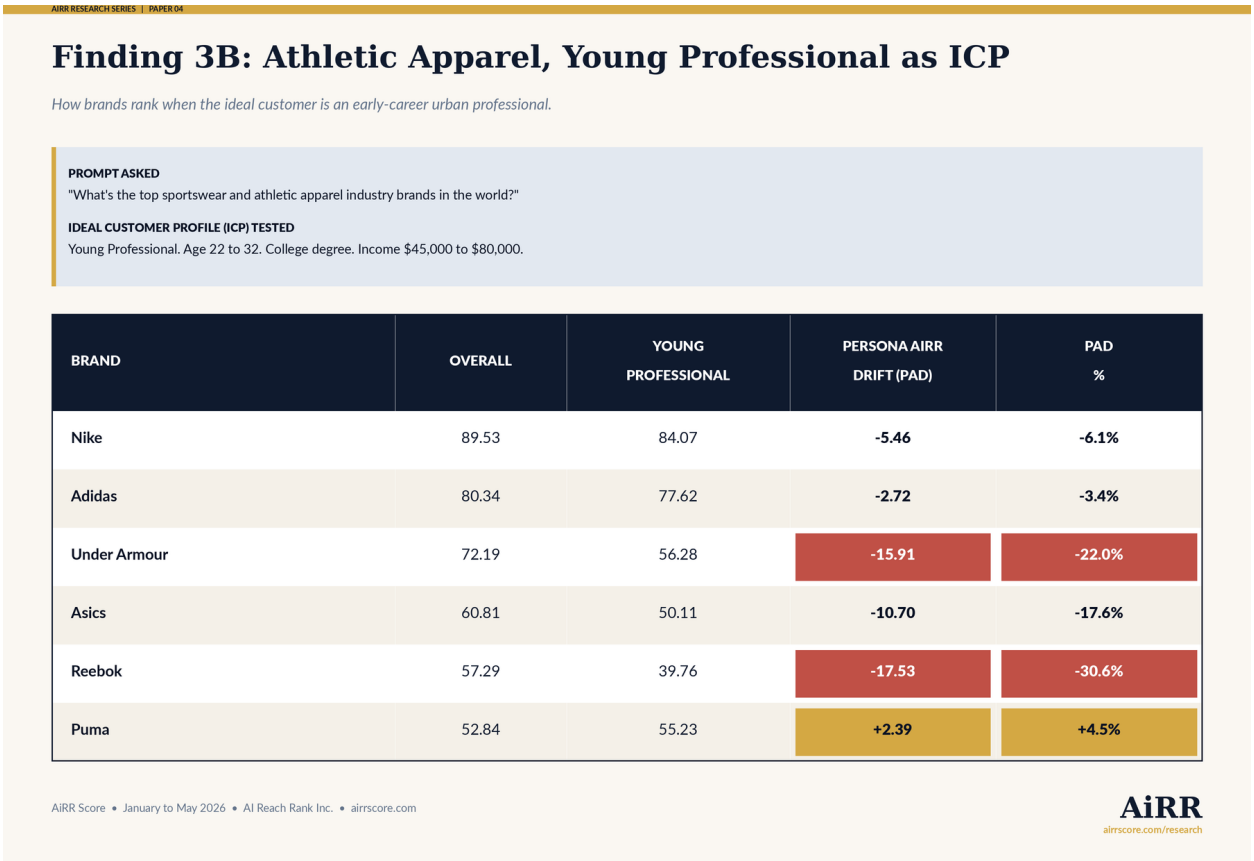
AiRR Score • January to May 2026 • AI Reach Rank Inc. • airrscore.com

AiRR
airrscore.com/research

Finding 3A: Athletic Apparel, Athletic & Active as ICP

Brand	Overall	Athletic & Active	PAD Points	PAD %
Nike	89.53	88.24	-1.29	-1.4%
Adidas	80.34	77.15	-3.19	-4.0%
Under Armour	72.19	60.23	-11.96	-16.6%
Asics	60.81	62.99	**+2.18**	**+3.6%**
Reebok	57.29	20.89	**-36.40**	**-63.5%**
Puma	52.84	34.73	**-18.11**	**-34.3%**

Finding 3B: If the ICP is Young Professional



Finding 3B: Athletic Apparel, Young Professional as ICP

Brand	Overall	Young Professional	PAD Points	PAD %
Nike	89.53	84.07	-5.46	-6.1%
Adidas	80.34	77.62	-2.72	-3.4%
Under Armour	72.19	56.28	**-15.91**	**-22.0%**
Asics	60.81	50.11	-10.70	-17.6%
Reebok	57.29	39.76	**-17.53**	**-30.6%**
Puma	52.84	55.23	**+2.39**	**+4.5%**

Interpretation. Reebok and Puma score in the mid 50s with a general audience. Under Athletic & Active conditioning, Reebok drops to -36.40 PAD (-63.5%) and Puma drops to -18.11 PAD (-34.3%). Asics moves in the opposite direction, gaining +2.18 PAD (+3.6%) because the informed audience recognizes it as a performance brand. Under Young Professional conditioning, every brand except Puma drops relative to overall, with Reebok (-17.53 / -30.6%) and Under Armour (-15.91 / -22.0%) falling hardest. Puma flips slightly positive at +2.39 (+4.5%), suggesting the lifestyle-leaning Young Professional persona reads Puma marginally more favorably than the general public does.

Implication. Reebok and Puma are present in the overall ranking and absent from the answer their actual category buyers see. The Young Professional and the Athletic & Active customer are not the same buyer, and ChatGPT does not recommend the same brand to both. Brand investment in mid-tier general awareness does not translate into mid-tier specialist recommendation.

FINDING 4: CYCLING, EXPERT AUDIENCES RESHAPE THE FIELD

Claim. In cycling, the overall AiRR Score overstates most brands' positions with both the Avid Cyclist and Luxury Seeker personas. The Luxury Seeker persona produces the largest swings, with Giant losing more

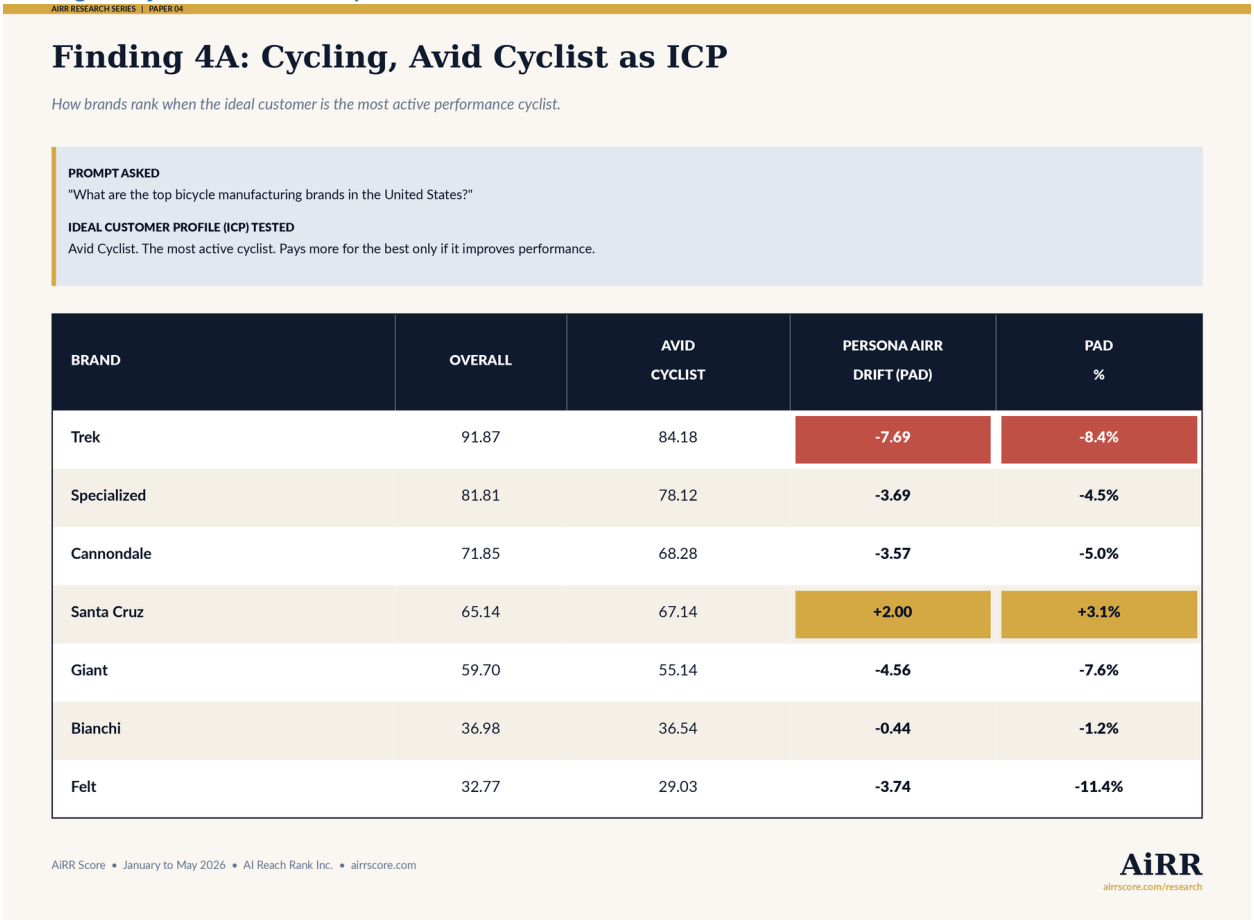
than 27 points (-46.6%) and Cannondale losing 13 points (-18.5%). Bianchi is the surprise positive mover under Luxury Seeker (+6.69 / +18.1%).

Prompt asked. *"What are the top bicycle manufacturing brands in the United States?"*

Personas measured.

- **Overall:** ChatGPT asked the prompt with no audience context.
- **Avid Cyclist:** "The most active cyclist who is willing to pay more for the best but only if it improves their performance."
- **Luxury Seeker:** "I am a high earner aged 30 to 55 with a household income over $150,000 per year. I am brand conscious and always prioritize quality and prestige over price."

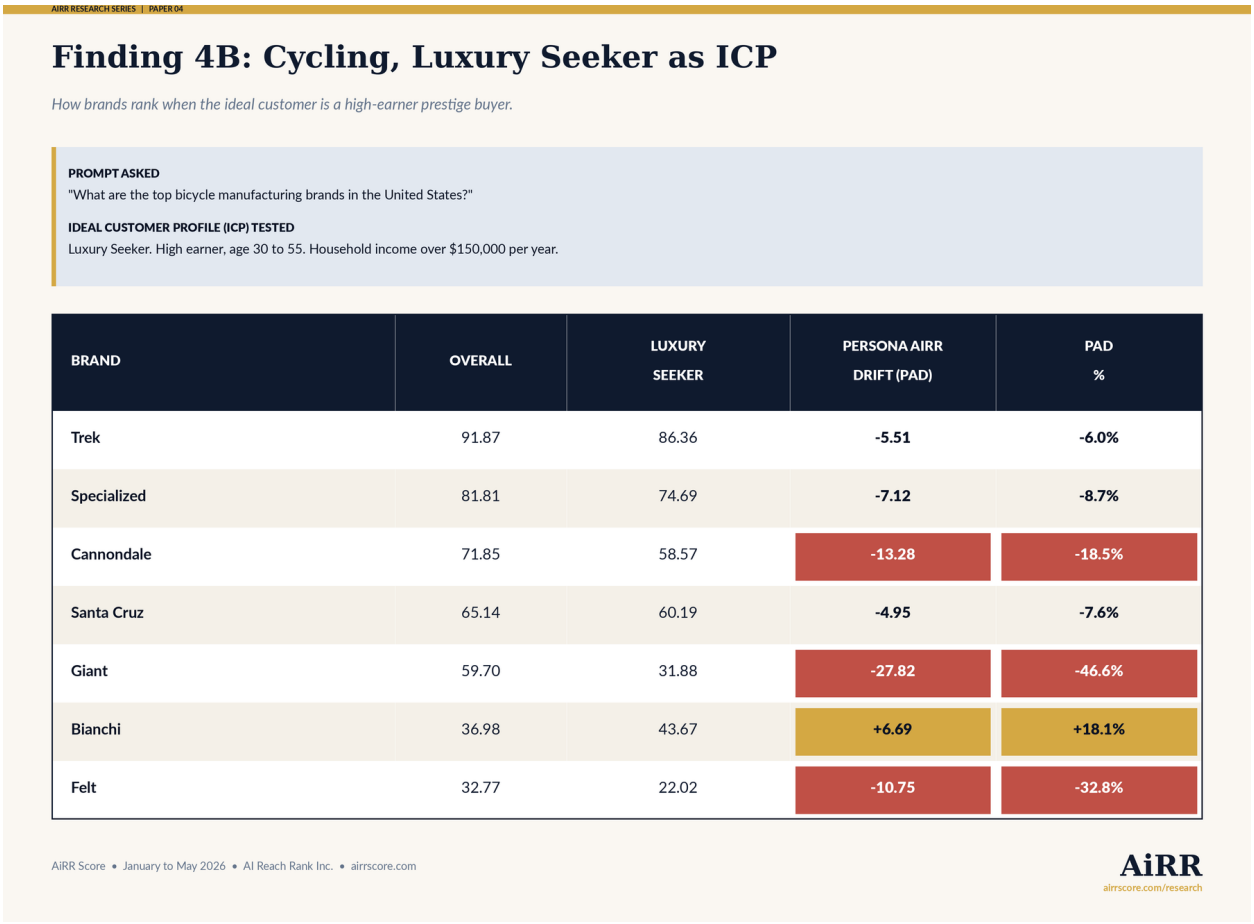
Finding 4A: If the ICP is Avid Cyclist



Finding 4A: Cycling, Avid Cyclist as ICP

Brand	Overall	Avid Cyclist	PAD Points	PAD %
Trek	91.87	84.18	**-7.69**	**-8.4%**
Specialized	81.81	78.12	-3.69	-4.5%
Cannondale	71.85	68.28	-3.57	-5.0%
Santa Cruz	65.14	67.14	**+2.00**	**+3.1%**
Giant	59.70	55.14	-4.56	-7.6%
Bianchi	36.98	36.54	-0.44	-1.2%
Felt	32.77	29.03	-3.74	-11.4%

Finding 4B: If the ICP is Luxury Seeker



Finding 4B: Cycling, Luxury Seeker as ICP

Brand	Overall	Luxury Seeker	PAD Points	PAD %
Trek	91.87	86.36	-5.51	-6.0%
Specialized	81.81	74.69	-7.12	-8.7%
Cannondale	71.85	58.57	**-13.28**	**-18.5%**
Santa Cruz	65.14	60.19	-4.95	-7.6%
Giant	59.70	31.88	**-27.82**	**-46.6%**
Bianchi	36.98	43.67	**+6.69**	**+18.1%**
Felt	32.77	22.02	**-10.75**	**-32.8%**

Interpretation. Trek leads in absolute AiRR Score across all three personas, but its PAD goes slightly negative under both Avid Cyclist (-7.69 / -8.4%) and Luxury Seeker (-5.51 / -6.0%), suggesting both informed audiences see Trek a little less favorably than the general public does. Specialized, Cannondale, and Giant all lose ground under Luxury Seeker conditioning, with Giant losing 27.82 points (-46.6%), the largest negative cycling PAD in the dataset. Bianchi inverts the pattern, gaining +6.69 (+18.1%) under Luxury Seeker even though its overall score is low. The Avid Cyclist persona produces only small movements across all seven brands, indicating that ChatGPT's recommendations for performance cyclists track close to its general recommendations.

Implication. Cannondale and Giant are credible cycling brands with enthusiast followings, but the Luxury Seeker persona does not see them. Bianchi's positive PAD under Luxury Seeker suggests the model associates the brand with prestige even though its general visibility is low. The persona conditions whether a brand's positioning is read as performance, prestige, or neither.

FINDING 5: TIME-SERIES MOVEMENT

Claim. Brand-level AI visibility moves at quarterly speed inside ChatGPT. 20-point composite shifts in 60 to 110 days are visible in this dataset and exceed the sensitivity of standard brand-health trackers.
Evidence. Each value below represents the change in the brand's Overall AiRR Score from the first day of measurement in this dataset to the most recent day of measurement. All values are in AiRR Score points on the 0 to 100 scale.
Largest composite declines (overall persona):

- Puma (Athletic Apparel): -21.94 in 104 days
- Bianchi (Cycling): -19.66 in 64 days
- Giant (Cycling): -14.27 in 64 days
- JetBlue (Airlines): -13.15 in 69 days

Largest composite gains (overall persona):

- Accenture (Professional Services): +14.78 in 55 days
- Felt (Cycling): +11.48 in 64 days

Interpretation. A 20-point composite move on a 0 to 100 scale within 60 to 110 days is a structural shift. Quarterly brand-health tracker sensitivity is typically narrower than this. The implication is that AI visibility decays and grows faster than traditional brand-health measurement detects.
Implication. Public-market analysts following brands like Puma can expect AI visibility to lead reported brand-health metrics by one to two quarters. The dataset is not yet long enough to confirm the lead-lag relationship empirically, but the magnitude and pace are consistent with that hypothesis.

---

4. The Future of Measurement: Brand × Prompt × Persona

The Persona Reversal findings point to a specific evolution in how AI search visibility will be measured. The first generation of AI visibility tools reports a single visibility score per brand. The data in this paper shows that number is an average across audiences the tool did not see. A brand can win the overall score and lose every persona that buys.
The second generation adds prompts. Tools now report visibility at the prompt level. That is closer to useful but incomplete. The same prompt produces different brand recommendations under different personas. Prompt-level visibility without persona conditioning is the same averaging problem at smaller granularity.
The future of measurement is the triple: **brand × prompt × persona.** The specific brand, the specific prompt, and the specific persona of the buyer asking. That combination is the unit of analysis where AI search recommendation actually happens. Anything less reports the average, and the PAD documented in this paper reaches 40.40 points in a single persona shift.
Why this matters more for CMOs than for any other reader. A Chief Marketing Officer running their own search on their own ChatGPT account is not seeing what their customers see. The ChatGPT account that returns a high ranking for the CMO's brand is automatically baking in the CMO's persona: their job history, prior searches, location, and inferred preferences. The model could be telling the CMO what the model thinks the CMO wants to hear. The actual buyer of the CMO's product is asking the same question from a completely different inferred persona, and is likely getting a different brand list.
Consider a concrete case from the data in this paper. McKinsey holds an overall AiRR Score of 78.90 in professional services, ranked #2 of 9 firms in the overall persona. Its CMO opens ChatGPT, runs the category prompt, sees McKinsey near the top, and concludes McKinsey is competitive. McKinsey's ICP for its enterprise innovation practice is a Big 4 partner commercializing frontier tech, asking the question from inside their own firm. That persona scores McKinsey at 51.17, ranked #6. The PAD is -27.73 points (-35.1%). ChatGPT typically returns three to five brands per recommendation answer. The actual buyer

sees four brands. McKinsey is not one of them. The CMO believes McKinsey is winning. The customer never sees the brand. **McKinsey is getting zero AI-influenced revenue from its target buyer while the marketing team is reporting #2 in the category.** The lesson generalizes: a CMO running a generic search from outside their company gets a different answer than a buyer running the same search from inside their company.

This is why the persona overlay is the most important layer of measurement, not an optional one. Any AI visibility platform that produces a single brand-level score, or even a brand-prompt score without persona context, is producing what amounts to a tainted number. The number can be correct in isolation and still be misleading as a strategic signal.

The AiRR Score reported in this paper is built on the triple. Each of the four Ps (Perception, Persistence, Presence, Prestige) is computable at the brand, prompt, and persona level. The composite shown in the tables is a summary of the underlying matrix. The measurement that matters for strategy is the matrix.

---

5. Implications

For marketers. The aggregate AiRR Score is a starting point, not an endpoint. Persona-level breakdowns reveal a different ChatGPT ranking for every customer segment a brand serves. The marketing team should read the PAD column for the persona that matches their declared ICP and treat that value as the actionable number. A negative PAD on the ICP persona means the brand is being oversold by overall visibility and the marketing team is operating on false confidence. A positive PAD on the ICP persona means the brand is being undersold by overall visibility and the brand is winning the audience that actually buys. Otherwise the optimization target is an average that does not match any real customer.

PAD is also a competitive metric. A brand can read its own PAD on its declared ICP alongside competitors' PAD on the same persona. A brand winning the ICP with +20 PAD while the top three competitors are all negative is the cleanest competitive signal the AI search layer produces.

For investors and analysts. Brand visibility decay inside AI search appears to run faster than traditional brand-health indicators detect. Puma's 22-point decline over 104 days, the steepest in the dataset, is the kind of signal that public-market analysts will eventually price in. The data infrastructure to price it does not yet exist outside independent measurement layers like AiRR.

For the GEO and AEO category. Tools that report a single brand-level visibility score are averaging across personas they do not measure. Recent industry analysis has flagged that LLM tracking tools face an accuracy crisis precisely because of personalization features [9]. The fix is not abandoning measurement. It is measuring the structural variables (persona, prompt context, and time) that operators need to act on.

For LLM providers. Personalization creates a measurement vacuum at the brand level. Operators can no longer answer "how am I doing in AI search" without persona context. Independent measurement at the persona level is the path through.

Quantified economic exposure. Paper 03 of the AiRR Research Series, *What a Position is Worth* [12], values Position 1 in a category-defining AI search query at $541,719 in annual influenced revenue (Total Annual AI-Influenced Revenue, or TAAIR), falling to effectively zero by Position 5. A brand holding Position 1 in the aggregate while sitting at Position 5 in the persona where its actual customers live is losing the full TAAIR delta on that segment. For an eight-segment business with persona-reversal exposure across half its segments, unrecognized revenue at risk runs into seven figures annually.

---

6. The AiRR Persona Reversal Index

The data in this paper is the inaugural release of an ongoing publication. The AiRR Persona Reversal Index, derived from the methodology in Section 2, will be updated on a recurring basis as the dataset expands across additional models, geographies, and industries. The index is intended for citation, reuse, and reproduction by independent journalists, analysts, and academic researchers. Methodology disclosures, raw data samples, and column definitions are available on request from steven@airrscore.com.

---

7. Conclusion

The most-cited measurement in marketing is about to change.

For two decades the dominant question of brand search ranking was showing up on the first page of Google on a list of ten blue links. Marketing teams optimized against it, paid for it, and reported it to their boards. The 0 to 100 AiRR Score, as a standalone aggregate, is the AI-search equivalent of that ranking. It is useful as a summary. It is dangerous as a strategy.

The data in this paper documents Persona AiRR Drift (PAD) of up to 40.40 points within a single brand when the audience persona shifts. Leaderboards invert at the top in professional services. Lifestyle athletic brands collapse with informed buyers. Cycling brands lose ground when the audience becomes a prestige buyer. A single overall AiRR Score reports the average of audiences the brand does not see, and the PAD can be off by more than 40 points on a 0 to 100 scale.

The unit of analysis must change. AI search recommendation is not a brand property. It is a property of the brand, the prompt, and the audience persona. The three together. Persona AiRR Drift is the measurement that makes the persona layer visible to marketing teams who would otherwise be operating on false confidence.

Three things follow.

For marketing leaders, the operating question changes from “what is my AiRR Score” to “what is my AiRR Score for my Ideal Customer Profile, and what is the PAD between the two.” A brand that wins the aggregate and loses its ICP is investing brand resources in an audience that is not paying it back. The PAD column is the audit.

For the GEO and AEO tooling categories, independent persona-level measurement is the missing layer. The tools that have shipped report visibility at the brand level or, more recently, at the prompt level. Neither layer answers the question this paper poses. Until persona-level measurement exists at scale, brands will continue to optimize against numbers that misrepresent the audiences they are trying to win.

For independent measurement to mean anything, it has to be willing to publish what brands do not want to see. KPMG winning a buyer that the marketing team at McKinsey does not yet know it has lost. Frontier emerging in a persona where the overall score made it invisible. Puma collapsing on the Prestige dimension while its lifestyle marketing continues. These observations are uncomfortable for the brands named. They are also where the value of the measurement lives.

This is another paper in AiRR’s ongoing measurement program. The findings here will be tested across additional models, additional geographies, and additional industries. The hypothesis is that the Persona Reversal will hold and that the magnitude will, if anything, grow. The future of marketing measurement is brand × prompt × persona, and the marketing teams that operate on that triple first will compound the advantage.

What’s your AiRR score?

References

[1] OpenAI. “Memory and new controls for ChatGPT.” 2024-2025 updates. https://openai.com/index/memory-and-new-controls-for-chatgpt/
[2] TechCrunch. “ChatGPT will now use its ‘memory’ to personalize web searches.” April 18, 2025. https://techcrunch.com/2025/04/18/chatgpt-will-now-use-its-memory-to-personalize-web-searches/
[3] Engadget. “ChatGPT’s new default model is more factual and better at personalization.” 2025. https://www.engadget.com/2165298/chatgpts-new-default-model-is-more-factual-and-better-at-personalization/
[4] “LLMs Have Rhythm: Fingerprinting Large Language Models Using Inter-Token Times and Network Traffic Analysis.” arXiv:2502.20589. https://arxiv.org/html/2502.20589v1
[5] Portkey. “Geo-location based LLM routing: Why it matters and how to do it right.” https://portkey.ai/blog/geo-location-based-llm-routing/
[6] “Understanding Biases in ChatGPT-based Recommender Systems: Provider Fairness, Temporal Stability, and Recency.” arXiv:2401.10545. https://arxiv.org/abs/2401.10545
[7] Wegner, D. M. (1987). “Paradoxical effects of thought suppression.” Journal of Personality and Social Psychology. Reviewed in: American Psychological Association. “Suppressing the ‘white bears.’” Monitor on Psychology, October 2011. https://www.apa.org/monitor/2011/10/unwanted-thoughts
[8] Nasr, M., Carlini, N., et al. “Scalable Extraction of Training Data from (Production) Language Models.” arXiv:2311.17035. 2023. https://arxiv.org/abs/2311.17035
[9] PPC Land. “LLM tracking tools face accuracy crisis from personalization features.” 2025. https://ppc.land/llm-tracking-tools-face-accuracy-crisis-from-personalization-features/
[10] MIT News. “Personalization features can make LLMs more agreeable.” February 2026. https://news.mit.edu/2026/personalization-features-can-make-llms-more-agreeable-0218
[11] “Personalized LLM for Generating Customized Responses to the Same Query from Different Users.” arXiv:2412.11736. https://arxiv.org/abs/2412.11736
[12] Perlman, S. (2026). *What a Position is Worth.* AiRR Research Series, Paper 03. AI Reach Rank Inc. airrscore.com/research/what-a-position-is-worth
Additional supporting research:
- Carlini, N. et al. “Extracting Training Data from Large Language Models.” USENIX Security 2021. https://www.usenix.org/system/files/sec21-carlini-extracting.pdf
- Carlini, N. et al. “Quantifying Memorization Across Neural Language Models.” ICLR 2023. https://arxiv.org/abs/2202.07646
- Search Engine Land. “What repeated ChatGPT runs reveal about brand visibility.” https://searchengineland.com/repeated-chatgpt-runs-brand-visibility-468552

About This Paper

Paper 04 of the AiRR Research Series.
How to cite. Perlman, S. (2026). *The Persona Reversal: Why AI Brand Visibility Is Not About Your Brand.* AiRR Research Series, Paper 04. AI Reach Rank Inc. airrscore.com/research/persona-reversal
Data verification. Raw scores, methodology disclosures, and column definitions available on request from steven@airrscore.com.
About AiRR. AiRR is the independent measurement standard for brand visibility in AI search. The AiRR Score is a 0 to 100 composite across four dimensions (Perception, Persistence, Presence, Prestige), measurable at the brand, prompt, and persona level. AiRR is operated by AI Reach Rank Inc., a Delaware C-Corporation headquartered in New York, NY.
Learn more: airrscore.com. **Research series:** airrscore.com/research