RANKFOR® INDEX 2026 · NORDIC-BALTIC + CEE EDITION

Nordic-Baltic + CEE AI Reputation Index

Methodology whitepaper for the Rankfor® Index 2026 ranking (ARI v3)

Version 3.0 · Published May 2026
Author: **Dmitrij Żatuchin**, PhD · Rankfor.AI

01 Purpose and scope

The Rankfor® Index is a composite measurement of how AI models describe consumer and enterprise brands across Northern, Baltic, and Central European languages. This whitepaper documents the methodology used to produce the Rankfor® Index 2026 (Nordic-Baltic + CEE Edition) ranking, based on the Cross-Language AI Reputation study conducted in April and May 2026.

The v3 index measures five dimensions:

1. Sentiment of AI-generated brand narratives (response-level)
2. Recommendation share in buyer-intent prompts
3. Quality of citation sources AI models reference
4. Cross-language consistency of the brand narrative
5. Stability of AI responses across repeated queries (proxy in Phase 1)

Rankfor® Index is designed to answer one question: when AI models are asked about these brands across the 11 study markets and 12 study languages (Lithuanian, Latvian, Estonian, Finnish, Polish, Czech, Slovak, Swedish, Norwegian, Danish, German, and English), how positive, recommendation-worthy, well-sourced, and cross-language-consistent is the resulting narrative?

02 Dataset

The Rankfor® Index 2026 (Nordic-Baltic + CEE Edition) is based on the Cross-Language AI Reputation study (Rankfor.AI, 2026), which collected **35,640** grounded AI responses across:

- **66 brands**, six per country across 11 home markets (Lithuania, Latvia, Estonia, Finland, Sweden, Norway, Denmark, Germany, Poland, Czech Republic, Slovakia)
- **12 query languages** spanning four language families: Baltic (Lithuanian, Latvian), Uralic (Estonian, Finnish), Slavic (Polish, Czech, Slovak), and Germanic (Swedish, Norwegian, Danish, German, English-as-control)
- **3 frontier AI models** (GPT-5.4 grounded via web search, Gemini 3.1 Pro Preview grounded via Google Search, Perplexity Sonar Pro web-augmented)
- **18 prompts** per (brand × language × model) cell, spanning open reputation (A1–A3), source attribution (B1–B3), competitive positioning (C1–C3), regional context (E1–E3), AI visibility and media (F1–F3); stability probes (D1–D3) deferred to Phase 2

Collection size: 18 prompts × 3 models × 12 languages × 66 brands = 35,640 total responses. Temperature 0.1, max tokens 4,096–8,192 per model. Citation backbone merged with the prior CEE 2026 study yields 131,667 cleaned citations across 21,077 unique cited domains.

The full dataset, collection protocol, and reproducibility package have been submitted to Springer *Discover Artificial Intelligence* for peer review.

03 The five components (ARI v3)

3.1 Sentiment (25% weight)

> **Definition** `w = 0.25`
>
> Per-response sentiment scored by CardiffNLP's `xlm-roberta-base-sentiment-multilingual` model on a -1 to +1 scale, then averaged across the brand's responses and rescaled to 0–100.

Sentiment is the most direct signal of how positively AI describes a brand to a buyer. Across 35,640 responses, the cohort mean sentiment is essentially neutral (-0.010), with 4.89% of responses scoring below -0.05 (negative-leaning) and 4.10% above +0.05 (positive-leaning).

Rationale. Sentiment carries the highest variance of any single signal in the study and the most journalistically interpretable meaning. A 0.30-point sentiment swing corresponds to a visibly different brand description.

3.2 Recommendation Share (25% weight)

> **Definition** `w = 0.25`
>
> Percentage of buyer-intent prompts (categories F1 + E1) in which AI mentions the brand by name, rescaled to 0–100.

Buyer-intent prompts are formulated without naming the brand (e.g., "Best mobile banking app in Poland for young professionals" rather than "Tell me about mBank"). Each prompt is submitted to all three models in the brand's home language and English. Detection uses a brand-alias matching system that checks for the brand (or known alternate names) anywhere in the response, plus position within the response and rank-in-list when the response uses a numbered recommendation format.

Rationale. Recommendation share is the dimension that maps most directly to revenue. A brand can have perfect cross-language consistency yet rarely be recommended. The component carries the same weight as sentiment because together they capture the two halves of the buyer experience: *does AI describe me well* and *does AI mention me at all when it matters*.

3.3 Source Quality Score (20% weight)

> **Definition** `w = 0.20`
>
> Weighted average of the brand's citation source types, normalized to a 0–100 scale across 11 source-type categories.

Source-type weights, based on editorial authority and AI training-data persistence:

SOURCE TYPE	EXAMPLES	WEIGHT
Tier-1 news	Reuters, Bloomberg, AP, major newspapers	1.00
Academic	Universities, peer-reviewed journals	0.90
Government	Regulatory filings, official databases	0.85
Industry report	Analyst firms, trade publications	0.80
Review platform	G2, Trustpilot, Capterra	0.60
Financial directory	Stock exchanges, company registries	0.55
Brand-owned	Corporate website, press releases	0.50
Wikipedia	Encyclopedic aggregation	0.30
Social media	Reddit, LinkedIn, X	0.20
Other web	Unclassified web content	0.15
Implicit knowledge	Unattributable parametric recall	0.10

Citations are extracted using a dual-LLM coding pipeline: the primary model classifies each attribution, a second model validates, and disagreements are resolved by grounded-mode URL-level citations from Perplexity Sonar Pro and the resolved Gemini vertex grounding-redirect URLs.

Rationale. Not all AI citations carry equal weight. A brand whose AI narrative is built on Reuters and academic coverage has a more durable reputation than one built on Wikipedia stubs and social-media posts. The weights reflect editorial permanence and AI training-data persistence.

3.4 Cross-language Consistency (20% weight)

> **Definition** `w = 0.20`
>
> Mean BGE-M3 (1024-dim) embedding cosine similarity between the brand's response in a given language and its English-baseline response, averaged across 11 home-language pairs and rescaled to 0–100.

Embeddings are computed with the BAAI/bge-m3 model. Pairwise cosine similarity is computed across 196,000 language pairs in the dataset. A brand whose AI narrative is semantically identical

across English and its 11 study languages scores near 100. A brand whose narrative diverges (different facts, different positioning, different tone in each language) scores lower.

Rationale. If AI tells a different story per language, the brand has lost control of its narrative. Cross-language consistency is the diagnostic that surfaces the Bilingual Penalty (multinationals penalized at home, local champions boosted at home).

3.5 Stability (10% weight)

> **Definition** `w = 0.10`
>
> Inter-model agreement proxy. Phase 1 uses a uniform value of 80 for all brands; Phase 2 will replace this with measured per-cell stability across 5 repeated queries.

In Phase 1 (the v3 release), the dice-roll stability prompts (D1–D3) were deferred. The Phase 2 collection (D1–D3 × 5 iterations × 66 brands × 12 languages × 3 models = 35,640 additional responses) will replace the uniform-80 proxy with measured per-cell stability scores. With real stability data, cohort spread will widen further and the lowest-stability brands will be exposed.

Rationale. AI models have inherent randomness. A brand whose description varies wildly across iterations cannot be governed because there is no stable narrative. Stability is a prerequisite for reputation management. The 10% weight is intentionally low until Phase 2 data lands.

3.6 Why citation density is not in the formula

A natural question arises: should we include citation density (citations per AI response) as a scoring component? We measured this and chose to exclude it.

Across the Rankfor® Index 2026 dataset, the relationship between citation density and ARI score is bidirectional. Some highly-ranked brands have low density because AI cites less when it trusts its parametric knowledge. Some lower-ranked brands have higher density because AI hedges by cross-checking a fragmented narrative.

Conclusion. Citation density is published alongside each brand profile as a diagnostic indicator, but not in the composite score. Including it would create perverse incentives, where brands could improve their score by suppressing AI citations rather than building narrative coherence.

3.7 Why Coverage was dropped in v3

The CEE 2026 v1 index used a "Coverage" component (15% weight): the percentage of markets where cross-language divergence stayed below an aligned threshold. At 7 CEE languages this added signal beyond the consistency average. At the 12 languages of the Nordic-Baltic + CEE Edition, Coverage became less informative because the recommendation-share component absorbed most of its variance, and because the wider language set made the binary "aligned/not aligned" cut less stable.

In v3, Recommendation Share inherits Coverage's weight (the v1 weights of 30% Consistency / 30% Rec Share / 15% Coverage / 15% Source Quality / 10% Stability rebalanced to 25% Sentiment / 25%

Rec Share / 20% Source Quality / 20% Consistency / 10% Stability). Sentiment was promoted to a first-class component because it had been an implicit input to Consistency in v1 and warranted its own line.

04 The composite formula

RANKFOR® INDEX COMPOSITE SCORE (ARI V3)

$$\text{ARI v3} = 0.25 \times \text{Sentiment} + 0.25 \times \text{Rec. Share} + 0.20 \times \text{Source Quality} + 0.20 \times \text{Cross-language Consistency} + 0.10 \times \text{Stability}$$

Result is on a 0 to 100 scale where higher values indicate more positively-described, recommendation-worthy, well-sourced, and cross-language-consistent AI reputation. In the 66-brand cohort, ARI v3 ranges from 42.86 to 66.51 with a mean of 53.71 and a standard deviation of 10.47. **18 brands score above 60**, 24 above 55, and 41 above 50.

Weight rationale

- **Sentiment (25%).** How positively AI describes the brand. The most direct signal of buyer-facing tone, with the highest cross-brand variance of any single component.
- **Recommendation Share (25%).** Whether AI mentions the brand when a buyer asks a category question. The metric that maps directly to revenue.
- **Source Quality (20%).** A proxy for narrative durability. AI trained on high-quality sources is harder to unseat.
- **Cross-language Consistency (20%).** Whether AI tells the same story in every language. Low consistency surfaces the Bilingual Penalty.
- **Stability (10%).** A noise floor (uniform-80 proxy in Phase 1; measured in Phase 2). Below a stability threshold, brand reputation is not governable.

v1 (CEE 2026) → v3 (Nordic-Baltic + CEE) change log. The v1 weights (30% Consistency, 30% Rec Share, 15% Coverage, 15% Source Quality, 10% Stability) produced a tight 34.4–56.3 band (std 2.04) because XLM-RoBERTa sentiment clusters near neutral for most brands and Coverage at 12 languages added little signal beyond the consistency average. v3 promotes Sentiment to a first-class 25% component, retains Recommendation Share at 25%, lifts Source Quality and Consistency to 20% each, and drops Coverage. Result: cohort spread widens to 34.7–69.5 (std 10.47), roughly five times the v1 band.

Source-quality weights were also updated to align with measured citation distributions. Tier-1 news restored to 1.00 (highest editorial weight); academic 0.90; government 0.85; industry report 0.80; brand-owned reduced to 0.50 (corporate content is influential but self-serving); Wikipedia 0.30; social 0.20; implicit knowledge 0.10.

05 Interpretation

60 - 70	**Leaders Club** AI-default brands with strong recommendation share, high source quality, and cross-language consistency. 18 brands in the cohort score in this band, led by Spotify (66.51).
50 - 60	**Strong Tier** Solid fundamentals with one or more weaker components. 23 brands score in this band, often held back by inconsistent cross-language narratives or contested category recommendation share.
42 - 50	**Development Tier** 25 brands. Significant opportunity for AI reputation work. The Bilingual Penalty (home-language underperformance vs English) is the most common diagnosis in this band.
Below 42	**Not observed** No brand in the Nordic-Baltic + CEE cohort scored below ARI 42.86 (cohort floor).

Rankfor® Index is not a brand quality ranking. It measures one specific thing: how AI describes the brand across the 12 study languages, how often AI recommends the brand at buyer-intent moments, and how durable the underlying source material is. A brand can rank low in Rankfor® Index and still be a great company; conversely, a high Rankfor® Index score does not guarantee business success. **Rankfor® Index is a diagnostic for AI reputation management, not a value judgment.**

06 Legal and ethical framework

Peer review

The underlying dataset and methodology have been submitted to Springer *Discover Artificial Intelligence*. This whitepaper documents the exact transformations used to produce the Rankfor® Index score.

Right of reply

Every brand included in the Rankfor® Index 2026 index has been offered the opportunity to submit a response. Responses will be published alongside the brand's score in the digital version at open.rankfor.ai/index-2026.

Reproducibility

The formula, weights, and source-type classification rules are published in this document. The underlying data is available to academic researchers on request.

Observation snapshot

Rankfor® Index scores reflect AI behavior at the time of measurement (April 2026). AI models, training data, and brand content change continuously. Scores are not predictive of future AI behavior.

Scope limitations

This edition measures 66 brands across 11 home markets, 6 brands per country. Exclusion from the Rankfor® Index 2026 is not a judgment on the brand or its AI reputation; it simply means the brand was not in the source study's sampling frame. Future editions will expand brand coverage.

No accusation language

Rankfor® Index is a composite score, not a claim about brand behavior. A low score means "AI currently describes this brand less positively, less often as the category default, or less consistently across languages than the cohort average," not "this brand misleads consumers."

07 How to cite

Żatuchin, D. (2026). **The Rankfor® Index 2026: Nordic-Baltic + CEE Edition.** Rankfor.AI Research Report. Submitted to Springer *Discover Artificial Intelligence*. Retrieved from open.rankfor.ai/index-2026

08 Contact

Dmitrij Żatuchin, PhD

dmitrij@rankfor.ai

Rankfor.AI · open.rankfor.ai

LinkedIn: linkedin.com/in/dizhat