Interpretability Engine Optimization (IEO): The Missing Optimization Layer in AI Commerce

Juan Carlos López Castaño
ORCID: 0009-0008-7471-0650
NSOLVIA Research

Version 1.2 · 5 September 2026
DOI: 10.5281/zenodo.22410276
Version 1.1 DOI: 10.5281/zenodo.22216684
Version 1.0 DOI: 10.5281/zenodo.22104281
Concept DOI (always resolves to latest): 10.5281/zenodo.22104280

Suggested citation: López Castaño, J. C. (2026). *Interpretability Engine Optimization (IEO): The Missing Optimization Layer in AI Commerce*. NSOLVIA Research. Version 1.2. DOI: 10.5281/zenodo.22410276

Abstract

Commerce is increasingly mediated by machines. A machine may retrieve an offer it does not fully understand; it cannot reliably constrain, compare, recommend, or transact on that offer without sufficient interpretation. Existing optimization disciplines — SEO, AEO, GEO and related AI optimization practices — address whether a business is found, extracted, cited, or acted upon. We did not identify an established discipline whose primary object is the question that precedes them all: **can the machine correctly interpret what the business sells?** This paper proposes **Interpretability Engine Optimization (IEO)**: the practice of structuring, resolving, enriching, validating, and representing commerce data so AI systems can correctly interpret offers before retrieval, comparison, recommendation, or action occurs. We define the discipline, its three objects of interpretation (product, brand, transaction), its core operations, a boundary test that separates it from neighboring practices, and a measurement approach implemented in a working instrument. We report empirical evidence from an operational audit corpus showing that interpretability failures are common within the audited corpus, measurable, and correctable — and that their correction produces large projected improvements in machine-facing readiness. IEO is proposed as an open discipline: anyone can practice it, with any tooling. IEO does not propose a new data format, protocol, model, or enrichment technique. Its proposal is the discipline itself: treating machine interpretability of source business data as an independent optimization objective that can be measured, improved, and tested. *Semantic discovery is downstream of interpretation.*

Keywords: Interpretability Engine Optimization · IEO · commerce interpretability · AI commerce · agentic commerce · machine interpretability · semantic commerce · structured product data · interpretability

---

1. The problem that precedes every other problem

In 2026, major commerce platforms connected merchant catalogs to AI shopping surfaces — conversational assistants, answer engines, and autonomous agents with checkout capability. These systems do not read product pages the way humans do. They filter, compare, and decide on **structured, machine-facing data**: official taxonomy categories, standardized attributes, machine-readable descriptions, explicit policies.

Yet large numbers of merchant catalogs entered this environment with machine-facing attributes incomplete, inferred, unconfirmed, or absent. The information usually *exists* — written by hand, for people, trapped in prose — but it is not *interpretable*: a machine cannot safely resolve what the product is, who stands behind it, or under what rules it can be bought.

The consequence is not a lower ranking. It is worse, and it is threefold:

1. **The offer may never enter the candidate set:** a product whose machine-facing fields cannot be resolved does not lose the comparison — it may never become eligible for it.
2. **The wrong recommendation:** where data is ambiguous, generative systems may infer or guess — and an inferred material, size behavior, or use case is a recommendation error delivered with confidence.
3. **The return that destroys the margin:** purchases made on misinterpreted data can result in avoidable returns — double freight, handling, and a disappointed customer.

Misinterpretation can create cost at three stages: eligibility, decision, and transaction.

A fourth effect compounds behind these three and is harder to quantify: **reputation**. A misinterpreted offer damages trust in two directions — with the buyer who received something other than what was understood, and plausibly with the recommending system itself, which staked its own reliability on a source that proved unreliable. Where downstream systems incorporate source reliability or historical performance, repeated interpretation failures could plausibly reduce the weight assigned to that source. This effect is stated here as a mechanism, not as a measured outcome.

Every downstream machine-mediated decision — search, answers, citations, agent actions — is constrained by the quality of the representation available to it. **If the machine misinterprets the product, downstream decisions may inherit that error.**

2. The proposal

NSOLVIA proposes Interpretability Engine Optimization (IEO) as a distinct optimization discipline.

IEO (general): *the practice of making a business's offers, entities, attributes, policies, and source facts machine-interpretable before downstream retrieval, recommendation, comparison, or action occurs.*

Commerce IEO: *the application of IEO to products, catalogs, brands and commerce operations.* (The scope of this paper and of the evidence reported.)

For technical audiences, an operational formulation: *making a merchant's offer sufficiently resolved, bounded and verifiable for AI systems to identify, constrain, compare, verify, cite, recommend, and transact on it — without inventing missing facts*. The unit of analysis is the **offer interpretation** — not the URL, not the SKU, but what a machine can correctly understand about the complete offer.

A note on the word "engine": IEO optimizes **source data for interpretation engines** — search engines, answer engines, generative systems, commerce platforms, and autonomous agents — never the engines themselves. We use "interpretability engine" as a collective term for any machine system that consumes business information and must form a usable interpretation of it.

The ladder. Each existing discipline answers a real question:

Discipline	Question	Functional center of gravity
IEO	Can machines correctly **interpret** what you sell?	UNDERSTAND
SEO	Can they **find** you?	FIND
AEO	Can they **extract** you?	ANSWER
GEO	Can they **surface and cite** you?	CITE / SURFACE
Agentic Readiness	Can they **compare, decide, act** on you?	ACT

IEO does not replace SEO, AEO, GEO or related AI optimization practices. It addresses the interpretability layer those downstream disciplines depend on. These are functional centers of gravity, not impermeable borders.

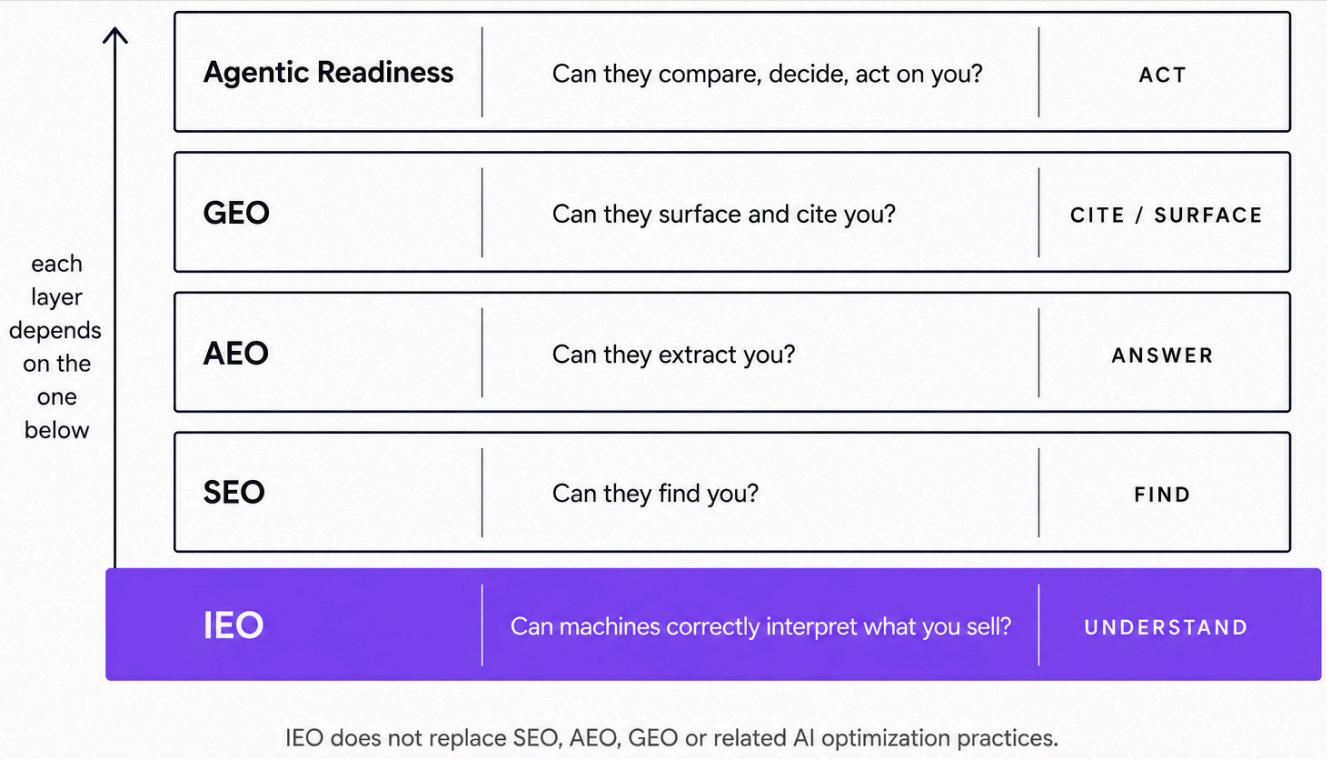

Figure 1. *The IEO ladder: interpretability as the layer preceding discovery, extraction, citation and action.*

3. The three objects of interpretation

Commerce interpretation is not only product interpretation. Machines evaluating an offer must resolve three distinct objects:

1. **Product Interpretability** — can the machine correctly understand *what is being sold*: identity, taxonomy, attributes, materials, uses, functional intent, purchase signals, restrictions, and the differences between near-identical items?
2. **Brand Interpretability** — can it understand *who stands behind it*: identity, provenance, manufacturing facts, certifications, claims and the evidence supporting them, promises the brand can sustain?
3. **Transactional Interpretability** — can it understand *under what rules it can be bought*: price, availability, shipping, returns, eligibility, exclusions?

Product tells the machine what is being sold. Brand tells it who stands behind it. Transaction tells it under what rules it can be bought. IEO makes all three interpretable.

"Commerce Interpretability" serves as the umbrella term for machine understanding across all three objects.

IEO applies across product, brand, and transactional interpretation. This paper's strongest empirical evidence is currently at the product level, where interpretability can already be measured and remediated at scale; brand and transactional interpretation follow the same logic and are earlier in their measurement maturity.

4. What IEO is — and the Boundary Test

Related work and scope. IEO does not claim that normalization, enrichment, taxonomy resolution, grounding, structured data, or semantic modeling are new techniques, nor that the underlying ideas of machine-actionability, task-specific fitness, provenance, or evaluation against declared requirements originate here. Close precedents exist for individual components and partial combinations of the objective this paper proposes. The claim of IEO is therefore **methodological integration and scope, not priority over its constituent techniques or concepts.** Five bodies of work stand closest and are acknowledged as principal antecedents.

Ontology engineering and evaluation. Grüninger and Fox (1995) introduced competency questions — requirements expressed as questions a representation must be able to answer — as simultaneously the design specification and the evaluation criterion of an ontology, with an informal statement formalized afterwards. The approach remains central to the field (Alharbi, Tamma, Payne & de Berardinis, 2026). IEO's resolution requirements inherit the principle that requirements and evaluation are one artifact. The difference in scope is the object and the judge: that literature evaluates whether *an ontology* adequately represents a domain, typically by hand with domain experts; IEO's object may be any source representation — a page, a feed, a record — and the judge is the machine's task performance, measured at scale on live sources.

FAIR and machine-actionability. Wilkinson et al. (2016) defined machine-actionability as a continuum in which a digital object enables an autonomous agent to identify what the object is, determine whether it is useful within the context of the agent's current task, determine constraints on its use, and take appropriate action. Wilkinson et al. (2019) then supplied an automated assessment architecture — Maturity Indicators, Compliance Tests and an Evaluator that reports what a machine "sees" when it visits a resource — and described its indicators as establishing a "contract of expectations and capabilities" between a resource and an agent. Machine understanding, task context, use constraints, downstream action, and the notion of a machine–resource contract therefore predate this paper. FAIR's scope is scientific data in repositories with curated metadata; IEO's is commercial source data on public surfaces that were never designed for machine consumption.

Contextual data quality. Wang and Strong (1996) defined data quality as fitness for use by data consumers, establishing that quality is relative to use rather than an intrinsic property, and identified representational quality as a distinct category. Later work developed semantic data quality frameworks based on intended use, incorporating provenance and pre-analysis assessment of fitness (Razzaghi, Greenberg & Bailey, 2022). The question *"interpretable for what?"* and the practice of assessing before downstream use are inherited from this tradition; the consumer there is typically a human analyst, whereas IEO's consumer is a machine that must resolve and act without asking.

Semantic interoperability. ISO/IEC 30182 and clinical terminology standards define the requirement that systems exchange information and correctly interpret its meaning, for which meaning must be agreed, consistent and clearly expressed. The objective of correct machine interpretation of meaning is not new; IEO applies it to a source that has no exchange partner and no agreed vocabulary — a merchant's own catalog.

Adjacent practices. The Semantic Web represents meaning; PIM/MDM systems govern product and master data; structured data encodes information in machine-readable formats; knowledge graphs represent entities and relationships; SEO, AEO and GEO optimize downstream discovery, answer and generative surfaces. Each may serve the practice; none defines it.

In the literature reviewed for this paper, we did not identify an antecedent that combines these elements around the same object: automated measurement and remediation of machine interpretability at the public commercial source-data layer, for merchant offers. IEO proposes organizing existing and emerging techniques around that distinct measurable objective: the quality of machine interpretation at the source-data layer — whether the resulting source representation is sufficiently resolved for machines to interpret the offer correctly.

IEO organizes existing techniques around one measurable outcome. Its typical pipeline:

Normalization → Semantic Resolution → Enrichment → Validation/Grounding → Machine-Interpretable Representation

First resolve what the product and its context actually are; then enrich with explicit, derived and verified meaning; then validate that nothing unsupported was introduced; the result is an interpretable record. The process can be executed by humans, rules, AI systems, semantic engines, or any mixture. **IEO does not depend on a specific implementation.**

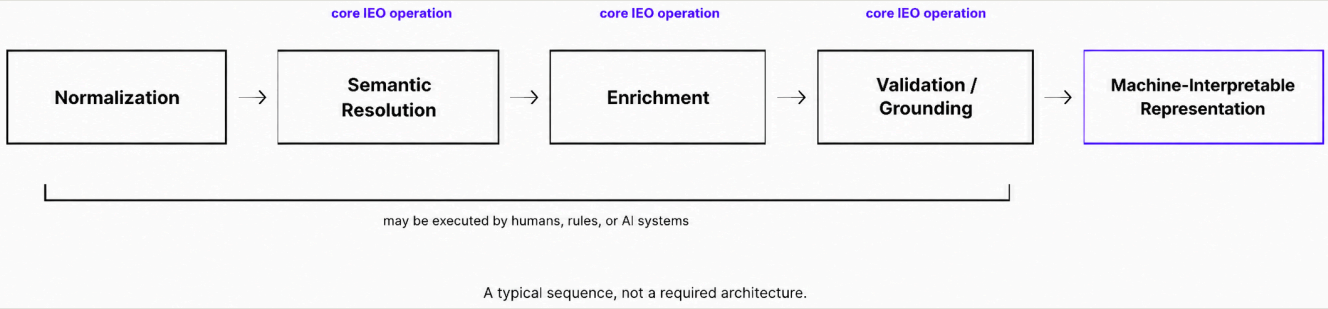

Figure 2. *The IEO pipeline. A typical sequence, not a required architecture.*

Because IEO borrows techniques from adjacent practices, a boundary test is necessary:

> **Boundary Test:** *Did the intervention improve the machine's understanding of the source — or did it only improve how the source is formatted, distributed, discovered, or surfaced?*

Layer	IEO?
Formatting (JSON-LD/XML/feeds as such)	Not by itself
Normalization	IEO operation
Semantic resolution	Core IEO operation
Enrichment	Core IEO operation
Validation / grounding	Core IEO operation
Exposure / distribution	Downstream deployment
Ranking / recommendation / action	Downstream outcomes

The practical form of the test: *if the same information were delivered through a different channel tomorrow, would the improvement in understanding still remain?* Interpretability survives the channel; positioning does not. And the proof that interpretability is a distinct property: **a record may be valid JSON-LD, perfectly normalized, and indexed — and still be semantically ambiguous. IEO evaluates the ambiguity that remains.**

A disambiguation for readers from machine learning: in ML research, "interpretability" (XAI) asks whether *humans can interpret the model.* Commerce IEO asks the reverse: whether *models can interpret the merchant.*

5. The artifacts

The output of IEO practice is a family of verified, machine-interpretable records:

- The **Semantic Product Record** — the product's verified meaning made explicit: resolved identity and taxonomy, normalized attributes, functional intents, use cases, purchase signals, safety text — each element traceable to merchant-supported evidence.
- The **Semantic Brand Record** — the brand's verified identity made explicit: who stands behind the offer, its provenance, manufacturing facts, claims and the evidence supporting them.
- The **Semantic Transaction Record** — the rules under which the offer can be bought: price, availability, shipping, returns, eligibility, exclusions. *(In current implementations, transactional rules are often captured within brand-level records; the conceptual separation stands regardless.)*

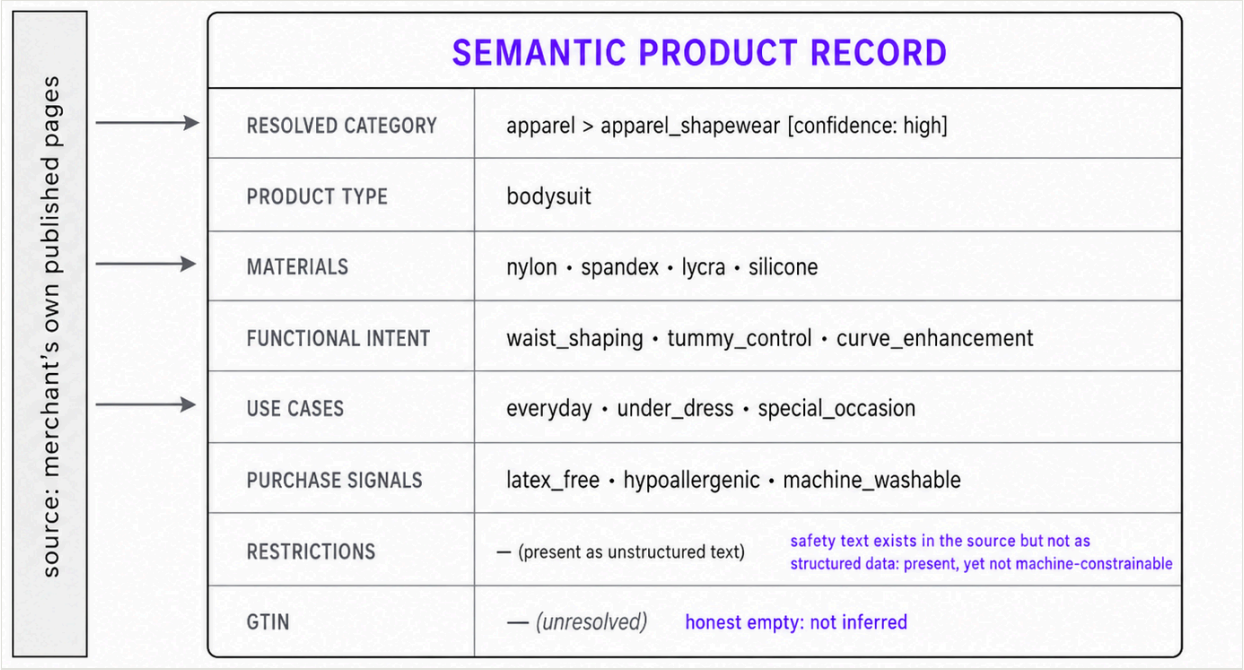

Figure 3. *A Semantic Product Record (PonteBella, ref. 30120), including two distinct kinds of gap: a field left unresolved rather than inferred, and a field whose content exists in the source as unstructured text and is therefore not machine-constrainable.*

Figure 4. *A Semantic Brand Record (PonteBella), including the transactional rules that determine whether an offer can be evaluated and purchased.*

Both figures use a brand owned by the author and operated as the research laboratory for this work; see Competing Interests.

The records exist in three states: the **source** (merchant data as it arrives), the **ingested record** (what the engine honestly resolved from an imperfect source), and the **gold record** (the verified, curated representation after remediation and — where the source is thin — human-confirmed facts). An ingested record must never be judged as if it were the gold record: it is the raw material on which the discipline demonstrates its value.

Two principles govern their construction. **Verified, never invented:** IEO does not prohibit inference; it prohibits *unsupported* inference. Resolving a product to its taxonomy category, or a stated composition to a normalized material value, is a semantic transformation entailed by evidence and is legitimate. Where evidence is insufficient, the field remains explicitly unresolved — an *honest empty* — rather than guessed. **The mirror principle:** human-facing and machine-facing layers state the same facts because they derive from the same source. The equivalence is factual, not literal: the machine-facing layer may carry identifiers, normalized values and provenance that no human reads on a page, but every factual assertion in it is traceable to human-verifiable evidence or an authoritative reference.

The records are the asset. The destinations — structured data, feeds, catalogs, agent files, APIs, protocols that do not exist yet — are adapters. *Protocols tell machines how to exchange commerce data. IEO asks whether the data being exchanged carries enough verified meaning to be correctly interpreted.*

6. Measurement

A discipline requires an instrument. Interpretability can be observed through resolution tasks — whether a machine can resolve identity, taxonomy, attributes, intent, use cases, trust/safety information, and retrieval language for a given offer. (We treat this as an initial observable framework, not a closed taxonomy; definitions are maintained in the public IEO glossary at nsolvia.com/ieo/glossary.)

In practice, NSOLVIA aggregates these observations into a working instrument — a catalog readiness audit producing a 0–100 score across three dimensions (structural, semantic, discoverability), plus a brand-level assessment. This audit is a broader commerce-readiness instrument: it includes interpretability-related structural and semantic measurements alongside downstream discoverability signals, and should not be read as a universal IEO metric. IEO itself does not prescribe a single score.

Three measurement terms follow, and the distinction between them is not cosmetic:

- the **Interpretability Baseline** — the observed current state, measured from the source representation on a versioned instrument that declares its reproducibility tolerance (zero for a de-

terministic instrument; its variance otherwise);

- the **Projected Interpretability Lift** — the estimated improvement represented by a remediated version, generated before deployment;
- the **Observed Interpretability Lift** — the difference measured after remediation has actually been deployed and the representation re-evaluated.

These three carry different evidential weight, and this paper is explicit about which is which. The baseline is a measurement. The projected lift is an estimate produced with generative assistance. The observed lift is the strongest form of evidence the discipline can produce, and it requires deployment. Section 7 reports baselines as measurements and lifts as projections, and labels them accordingly.

The discipline admits other instruments. Any tool that measures whether machines can correctly resolve an offer's meaning — by these or other observable tasks — is measuring commerce interpretability. That independence is what makes IEO a discipline rather than a product.

7. Empirical evidence

Corpus. This paper reports the **IEO Dataset V1**: 177 audit records, of which **159 produced a persisted report with scores**, covering **105 distinct products across 63 stores in 10 verticals** (fashion, jewelry, toys, food, home, health, pets, beauty, electronics, sports) on four commerce platforms. Aggregate findings below are reported over the analysis subset of **N = 155 scored records** measured end to end with a **single versioned instrument (`ss0@12a70bd`)**, which is what makes the reliability figures meaningful; four laboratory records measured with a later, verified-equivalent instrument version are tagged in the dataset and excluded from aggregate reporting. Each store contributes one product selected by the instrument as representative of the catalog and, where the platform allowed it, a second product taken arbitrarily from the same catalog — the contrast between the two is itself part of what the corpus measures.

Comparability rule. Any lift computed from two measurements requires that both share the same **dimensional maxima**. A total computed over different denominators is not the same total: comparing across a scale change can produce an apparent lift that is a pure artifact of the scale. Where the maxima differ, the earlier measurement is discarded as a baseline rather than rescaled. This rule is part of the dataset's inclusion criteria.

Correction to Version 1.0. Version 1.0 of this paper stated that the instrument's dimensional scales had been revised during the period covered by this work, with the semantic and discoverability maxima both increased. **That statement was incorrect and is withdrawn.** Verification against the complete audit record shows the dimensional maxima constant at (30, 40,

30) across all 493 scored rows recorded between May and August 2026, without a single exception, and constant across the 61 rows of the laboratory brand. The error originated in a misreading of a stored field, in which a projected post-remediation value was read as a denominator. The comparability rule itself is unaffected and stands as stated; what is withdrawn is the empirical claim that a scale change occurred in this corpus. The statement is retired here rather than removed, because a deposited paper does not delete what it said.

An earlier corpus (n = 80) was published in NSOLVIA's Semantic Commerce Layer whitepaper using a prior instrument version; this paper reports the IEO Dataset V1, measured end to end with a single versioned instrument, and the two should not be pooled.

Provenance of the instrument. The measurements reported here were not produced for this paper. The instrument — a catalog readiness audit published as a Pillar Document on 30 June 2026 — was already in operational use, and the corpus below accumulated through that use. The discipline described in this paper was formalized around an instrument that existed first, not the reverse. This is stated as a matter of method: it explains why the corpus is operational rather than designed, which is also the source of the selection bias declared in Section 10.

Ethics and scope of access. All measurements derive from publicly accessible pages. No authenticated or private data was accessed, nothing was written to any store, and no communication was sent to any merchant. Aggregate results are reported; individual audited stores are not identified.

Instrument reliability (test-retest). Because the instrument's product selection is deterministic when a store has a dominant category, re-running a store frequently returns the same product — producing, at no additional cost, 52 same-product measurement pairs. Across those pairs:

- the **measured baseline score was identical in 52 of 52 pairs (100%)**;
- **all fields were identical in 40 of 52 pairs (77%)**.

The residual variation appears in the *projected* post-remediation score, which is generated rather than measured.

The scope of this reliability claim should be stated precisely, because the distinction is the point. What is reproducible is the **measurement**: baseline score, dimensional breakdown and declared gaps are byte-identical across repeated runs. What is *not* reproducible in the same sense is the **generated representation** underlying the projection: on a laboratory product examined field by field across two consecutive runs, the enriched record diverged in 4 of 32 fields — not in the resolved category, functional intent or use cases, which were stable, but in the enumeration of materials and in the attribute schema itself, where the second run used different keys for the same

product. This is expected behaviour of a generative step and is precisely why this paper reports measurement and projection as different quantities. **The baseline is reproducible; the projection is an estimate.** A discipline that does not separate these two will eventually mistake one for the other.

Findings.

- **Interpretability failure is common within the audited corpus, and it is measurable.** Mean measured baselines by vertical ranged from **28.5 to 41.1 on a 0–100 scale** — that is, in every vertical examined, the average store's source representation left most of what a machine needs to interpret an offer unresolved.
- **Failures are observed across different levels of human-facing content quality.** Pages with rich, well-written human copy measured lower structural readiness than thin pages in the same run. Content richness and machine-facing readiness are distinct properties; the dominant pattern is not missing information but *uninterpretable* information.
- **Category resolvability varies widely and tracks taxonomy coverage, not store quality.** The share of products whose category could be resolved to a supported leaf ranged from **47% to 100%** across verticals. Where resolvability was lowest, projected improvement was also lowest — consistent with the interpretation that resolution, not decoration, is what drives readiness.
- **Correction produces large projected readiness lifts.** Projected post-remediation scores rose substantially across the corpus, with the smallest projected gains concentrated precisely in the verticals where category resolution failed most often. These are projections, not deployed outcomes, and are reported as such.
- **Structural presence is not structural quality — and the instrument measures presence.** Within the corpus, products whose fields were populated with weak or uninformative content scored close to products whose fields were populated with well-formed content: the structural dimension registers *whether* a field carries a value, not whether that value is interpretable. The bias this introduces runs in a conservative direction — the baselines reported above are, if anything, generous, and a stricter instrument would report lower. What presence-based measurement cannot see appears one layer down: the discrimination between a populated record and a *resolved* one is carried by the semantic dimension, which is also where the largest remediation margin is consistently observed.

External convergence. Independent work reaches a compatible conclusion. A 2026 preprint on agent-ready websites reports a controlled experiment across five tasks, three agent models and 300 runs, in which strict task success rose from **74 of 150 runs (49.3%) on the baseline site**

to 134 of 150 runs (89.3%) on the agent-ready site, with the largest gains in product-detail extraction, comparison, and multi-constraint selection (Elnaffar & Rashidi, 2026; preprint, not peer-reviewed). Notably, that framework names *agent interpretability* as its first dimension — an independent arrival at the same concept from a different starting point.

Falsifiability. The central claim is testable in two stages. **Stage 1 — remediation:** compare the offer as received against the same offer after IEO processing (audit → resolution → enrichment → validation), measuring what the machine can now correctly resolve. This demonstrates remediation capacity on real, imperfect sources. **Stage 2 — the IEO Counterfactual Product Test:** take one fully processed offer and produce a controlled degraded version (category genericized, attributes removed, application constraints deleted, everything else — price, images, brand — held constant). Give the same AI system the same questions and access to both representations, and measure attribute extraction accuracy, category resolution, constraint satisfaction, differentiation between near-identical items, hallucination rate, and task completion against a golden interpretation record. Protocol requirements: same model and version, same system configuration, same questions, same retrieval permissions and settings; representations evaluated separately and blinded — the model must never see both versions simultaneously; randomized order across multiple runs; gold labels defined in advance; results reported with confidence intervals. All material variables except representation quality are held constant; any difference in machine understanding is attributable to representation quality alone. If processed representations do not outperform, the discipline's premise fails. We invite replication.

8. The economic translation of interpretability

IEO produces a technical outcome first: improved machine interpretation. Its business value is established downstream, where that improvement can be connected to measurable merchant outcomes — reduced manual remediation, fewer interpretation-driven errors, more accurate product matching, lower support burden, fewer avoidable returns, or incremental qualified demand. These effects must be measured, not assumed; where estimates are made, they must derive from the merchant's own data.

The measurement path is a chain of observable bridges: **interpretability baseline → interpretability lift → operational effect → platform effect → commercial effect.** The discipline does not jump from the first bridge to the last; it measures each one. (Example: attribute ambiguity ↓ → correct matching ↑ → recommendation errors ↓ → information-driven returns ↓ → margin preserved.)

A methodological note favors this discipline: unlike downstream optimization practices — which measure against surfaces that shift beneath the measurement (model updates, retrieval changes, non-deterministic responses) — IEO controls its experimental object. The same offer in two representations, questioned by the same system under the same conditions, isolates representation quality as the variable. A discipline born with a controllable experiment provides a comparatively controlled basis for building an economic evidence base.

IEO identifies and corrects representation failures that can affect search, feeds, marketplaces and conversion today; autonomous agents may amplify those downstream consequences.

9. What IEO is not — and what it opens

IEO is not keyword optimization, not content generation, not visibility tracking, not schema markup as such, and not a product. **IEO is proposed as an open discipline.** NSOLVIA does not require its own tools to practice it: IEO can be practiced manually, with enterprise systems, or with specialized instruments of any provenance. NSOLVIA built a commerce-native stack to measure and automate it — and maintains this definition as a public reference.

Implications for practice. An open discipline creates work beyond any single vendor. A practitioner profile emerges around auditing, remediating and maintaining source interpretability — for merchants directly and through agencies. Toolmakers and engineers have room to build independent instruments: validators, benchmarks, counterfactual testers, interpretation simulators, and connectors linking interpretation quality to analytics and campaign performance — with any engine, including their own. The discipline does not require NSOLVIA's tools; it requires the measurable objective they share.

10. Limitations

This proposal is made with its current limits stated. The audit corpus is operational, not a randomized sample, and may carry selection bias toward stores reachable by the instrument. Measurements to date come from a single provider's instrument; the discipline invites independent instruments and replication.

Post-remediation figures reported here are projections, not deployed outcomes: observed interpretability lift requires deployment and re-measurement, and is reserved for a subsequent version of this work. The distinction between the two is not merely terminological, and we state the expected direction in advance: a first post-deployment measurement in the author's own laboratory brand indicates that **observed lift is substantially smaller than projected lift for the same product**. This is not evidence against the projection — the projected figure esti-

mates the achievable ceiling if every available remediation were applied, whereas any given deployment is necessarily a subset of it, and the dimensional breakdown shows the remaining margin concentrated in the semantic dimension. Reporting the two as if they were the same quantity would misstate both.

Empirical evidence is concentrated at the product level; brand and transactional interpretation follow the same logic but are earlier in measurement maturity. The instrument's structural dimension measures field presence rather than field quality: two products with equally complete but unequally interpretable content can receive similar structural scores. This biases reported baselines upward rather than downward, and later instrument versions should incorporate content-quality signals at the structural layer.

Economic effects are stated as measurable hypotheses, not demonstrated outcomes; longitudinal merchant studies are pending, and the reputational effect described in Section 1 remains a stated mechanism with no measurement attached. Platform claims are most extensively documented on Shopify, where category-level structured fields shipped platform-wide; behavior on other platforms is inferred from their public data surfaces.

11. Terminology

The following terms are used throughout. Full definitions are maintained in the public IEO glossary at nsolvia.com/ieo/glossary.

A note on the acronym. "IEO" is used in this work exclusively for **Interpretability Engine Optimization** as defined here. The acronym has other expansions in circulation, referring to different practices with different objects; this work refers only to the one defined above. The convention throughout is that the first mention in any section uses the full form, and the acronym is used thereafter for brevity.

IEO — Interpretability Engine Optimization. The practice of making a business's offers, entities, attributes, policies, and source facts machine-interpretable before downstream retrieval, recommendation, comparison, or action occurs.

Interpretability Engine. Any machine system that consumes business information and must form a usable interpretation of it — search engines, answer engines, generative systems, recommendation systems, commerce platforms, autonomous agents.

Commerce IEO. The application of IEO to products, catalogs, brands, offers and commerce operations.

Product Interpretability. The degree to which a machine can correctly understand what is being sold: identity, taxonomy, attributes, materials, use cases, functional intent, purchase signals, restrictions, and meaningful differences between related products.

Brand Interpretability. The degree to which a machine can correctly understand who stands behind an offer: identity, provenance, manufacturing facts, certifications, supported claims and trust-relevant facts.

Transactional Interpretability. The degree to which a machine can correctly understand the rules under which an offer can be evaluated or purchased: price, availability, shipping, returns, eligibility, exclusions.

Semantic Resolvability. The degree to which the meaning required for machine interpretation can be resolved from available evidence without unsupported inference.

Semantic Product Record. The product's verified meaning made explicit: resolved identity and taxonomy, normalized attributes, functional intent, use cases, purchase signals and safety information.

Interpretability Baseline / Projected Interpretability Lift / Observed Interpretability Lift. The measured current state; the estimated improvement of a remediated representation before deployment; and the difference measured after deployment and re-evaluation.

Interpretability Debt. The accumulated gap between what a business knows about its offers and what its machine-facing representations make explicit enough for machines to understand correctly.

Competing interests

The author is the founder of NSOLVIA, which develops commercial instruments for measuring and remediating commerce interpretability. The brand used in Figures 3 and 4 is owned by the author and operated as the research laboratory for this work. NSOLVIA proposes Interpretability Engine Optimization as an open discipline.

Acknowledgments

NSOLVIA Research.

References

Alharbi, R., Tamma, V., Payne, T. R., & de Berardinis, J. (2026). Use of competency questions in ontology engineering: A survey. *AI Magazine*. https://doi.org/10.1002/aaai.70054

Elnaffar, S., & Rashidi, F. (2026). *Designing Agent-Ready Websites for AI Web Agents: A Framework for Machine Readability, Actionability, and Decision Reliability*. arXiv preprint arXiv:2607.12056. [Preprint; not peer-reviewed.]

Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. In *Proceedings of the IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal.

ISO/IEC 30182:2017. *Smart city concept model — Guidance for establishing a model for data interoperability*. International Organization for Standardization.

López Castaño, J. C. (2026). *The Semantic Commerce Layer: A Framework for Machine-Readable Commerce Data*. NSOLVIA Research.

Razzaghi, H., Greenberg, J., & Bailey, L. C. (2022). Developing a systematic approach to assessing data quality in secondary use of clinical data based on intended use. *Learning Health Systems, 6*(1), e10264. https://doi.org/10.1002/lrh2.10264

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, M. D., Dumontier, M., Sansone, S.-A., et al. (2019). Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data, 6*, 174. https://doi.org/10.1038/s41597-019-0184-5

Version history

Version 1.2 (5 September 2026) — Related work and precision release. (1) Section 4 is expanded to acknowledge five principal antecedents — ontology engineering and competency questions (Grüninger & Fox, 1995; Alharbi et al., 2026), FAIR machine-actionability and its evaluation framework (Wilkinson et al., 2016, 2019), contextual data quality (Wang & Strong, 1996; Razzaghi et al., 2022), and semantic interoperability standards — and restates the claim of this paper as

methodological integration and scope rather than priority over constituent concepts. (2) Section 5's "verified, never invented" principle is clarified: inference is not prohibited; unsupported inference is. (3) The mirror principle is restated as factual rather than literal equivalence. (4) The Interpretability Baseline definition now requires the instrument to declare its reproducibility tolerance. (5) A declared limit of the counterfactual test design is added, with an ablation extension reserved for subsequent work. (6) Statements regarding trademark rights are removed: the discipline's openness is established by its licence and by the absence of any requirement to use NSOLVIA's tools, and the paper takes no position on trademark matters. No corpus figures, findings or reliability statistics change.

Version 1.1 (30 August 2026) — Corrective release. Five changes, none of which alters the corpus figures, the reliability statistics, the findings, the falsifiability protocol or the limitations. (1) A single empirical statement in Section 7 is withdrawn: Version 1.0 asserted that the instrument's dimensional scales had been revised during the period covered by this work. Verification against the complete audit record (493 scored rows, May–August 2026) shows the dimensional maxima constant at (30, 40, 30) throughout. The claim is retired in place, with the correction stated in Section 7. (2) The statement that the term "is not trademarked and no vendor holds exclusive rights to the practice" is replaced throughout by a statement of NSOLVIA's own position — that it proposes the discipline as open and does not claim trademark rights in it — since the original made an unverifiable assertion about third parties. (3) The Abstract's claim that no existing discipline addresses the question is restated as a finding of this review rather than an absolute. (4) Section 11 adds a note on the acronym: several expansions of "IEO" are in circulation, and this work refers only to the one defined here. (5) Section 7 adds a note on the provenance of the instrument, which predates the framework. The comparability rule, the corpus figures, the reliability statistics, the findings, the falsifiability protocol and the limitations are identical to Version 1.0.

Version 1.0 — Initial public proposal. Deposited 26 August 2026. DOI: 10.5281/zenodo.22104281.
