IEO Field Guide

A working guide for practitioners

NSOLVIA Research

Version 1.0 · 2026 · **License:** CC BY 4.0

Based on: *Interpretability Engine Optimization (IEO): The Missing Optimization Layer in AI Commerce.* DOI: 10.5281/zenodo.22104281

What this is

A working guide for anyone who audits, remediates or maintains the machine-interpretability of commerce data — agency strategists, consultants, catalog managers, merchants doing it themselves.

It is not a summary of the founding paper. It is the part you use on a Monday.

One assumption: you already accept the premise. If you want the argument, the evidence, or the falsifiability protocol, read the paper. If you want to know what to do, keep reading.

The one question

Every decision in this practice reduces to one 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?**

The practical form, and the one to use in a client meeting:

> *If the same information were delivered through a different channel tomorrow, would the improvement still remain?*

Interpretability survives the channel. Positioning does not.

Use it as a filter on your own scope of work. If a deliverable would not survive a channel change, it is a fine deliverable — it is simply not this work, and it should not be sold as this work.

The five steps

Normalization → Semantic Resolution → Enrichment → Validation → Representation

Step	What you are doing	Done when
1. Normalization	Making inconsistent data consistent. Same attribute, same expression, across products.	Two products of the same type express the same attribute the same way.
2. Semantic Resolution ★	Resolving what the product actually *is* against an official taxonomy.	The category resolves to a supported leaf, not to an internal label.
3. Enrichment ★	Making explicit what was implicit: intent, use cases, decision criteria, purchase signals.	A machine could argue for this product using only its structured data.
4. Validation ★	Confirming nothing unsupported was introduced.	Every value traces to something the merchant published or confirmed.
5. Representation	The interpretable record, ready for any destination.	—

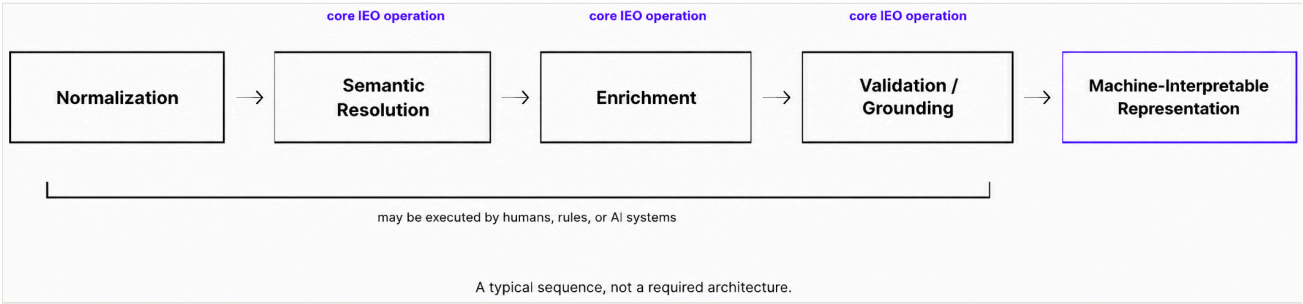

Figure 2. *The IEO pipeline. A typical sequence, not a required architecture.*

★ = core operation. If your engagement skips steps 2–4, you are doing formatting.

Order matters and it is not negotiable. Enriching before resolving produces detailed descriptions of a product nobody can categorize. Skipping validation produces a catalog you cannot defend.

The three objects — your audit scope

Do not scope an engagement to products only. Machines resolve three objects, and an offer fails if any of them is unresolvable.

PRODUCT — what is being sold
Identity · taxonomy · attributes · materials · use cases · functional intent · purchase signals · restrictions · differences between near-identical items

BRAND — who stands behind it
Identity · provenance · manufacturing facts · certifications · supported claims · trust-relevant facts

TRANSACTION — under what rules it can be bought
Price · availability · shipping · returns · eligibility · exclusions

> **The transaction object is the one almost everyone forgets.** An agent asked to *buy* — not recommend, buy — must resolve whether the store ships to that country, at what cost, in how many days, whether the item is returnable, within what window, and whether a refund is money or store credit. Perfect product data fails at the last step if those cannot be resolved.

Practical note: in most implementations today, transactional rules live inside brand-level records. That is fine. The conceptual separation stands regardless, and you should audit for it separately.

The four kinds of gap you will find

This is the most useful thing in this guide. When you audit a catalog, gaps are not all the same, and they do not get the same treatment.

> **1 MISSING**
>
> the information does not exist anywhere.
>
> *Example:* no GTIN, in the catalog or on the page.
> **Treatment:** leave it explicitly unresolved. Flag it with guidance on how the merchant can obtain it. **Never infer it.**

2 **PRESENT BUT UNSTRUCTURED**

the information exists, in prose, legible to any human, and cannot be acted on by a machine.

Example: safety guidance written as a paragraph, never expressed as a constraint.
Treatment: structure it. Keep the prose unchanged for humans. This is the highest-value gap in most catalogs, and the one merchants are most surprised by.

3 **AMBIGUOUS**

a value exists but does not resolve.

Example: category set to "SS-24 Collection."
Treatment: resolve against official taxonomy. This is usually the highest-leverage fix in the entire engagement.

4 **INCONSISTENT**

the same fact is stated differently in different places.

Example: return window says 30 days on one page, 15 on another.
Treatment: resolve with the merchant, then state once. Never pick one silently — inconsistency is a decision the merchant has to make, not you.

The rule that protects you

> ***Verified, never invented.*** *Where evidence is insufficient, the field stays explicitly unresolved — an* honest empty *— rather than inferred.*

This will feel wrong the first time a client asks why a field is blank. Say this:

> "We left it empty because your pages don't state it. Anything we put there would be our guess wearing your name — and if a customer received something different, it would be your liability, not ours. Here's what it would take to fill it properly."

An honest empty is a finding, not a failure. And it is the only kind of catalog you can defend when it is audited.

Its companion, the mirror principle: human-facing and machine-facing layers state the same facts, because they come from the same source. Anyone can verify the match by reading the page. Never write for machines what a human cannot see — that is the oldest manipulation in search, and modern systems penalize it.

Measuring honestly

Three terms. Using them loosely is the fastest way to lose credibility in this field.

Term	What it is	What you may say
Interpretability Baseline	The current state, **measured**	"Your catalog scores X today."
Projected Lift	The estimated improvement, **generated** before deployment	"If everything were applied, it would score Y."
Observed Lift	The difference **measured** after deployment and re-measurement	"We deployed in March, re-measured in May: X to Z."

Three rules that will keep you honest:

1. **Never report a projection in the vocabulary of a measurement.** "It would score Y" and "it scores Y" are different sentences and different promises.
2. **Expect the observed lift to be smaller than the projected one.** The projection estimates the ceiling if *everything* were applied; any real deployment is a subset. Tell the client this *before* you deploy, not after they ask.
3. **Comparability rule:** two measurements are comparable only if they share the same dimensional maxima. If the instrument's scale changed between the two runs, the difference is an artifact, not a lift. Discard the earlier baseline rather than rescaling it.

And know your instrument's bias. Most instruments measure whether a field is *populated*, not whether its content is *interpretable*. That biases scores upward — meaning the real gap is usually larger than the number says. Say so.

Self-assessment: ten questions

Run these against any catalog before quoting an engagement.

1. ☐ Does the category of a representative product resolve to an official taxonomy leaf, or to an internal label?
2. ☐ Are the attributes that category requires populated — and are the values interpretable, or just present?
3. ☐ Could a machine state who this product is *for*, and *when*, using only structured data?
4. ☐ Could it justify recommending this product in two sentences, using only structured data?
5. ☐ Could it distinguish this product from the two most similar products in the same catalog?
6. ☐ Are safety or usage restrictions expressed as constraints, or only as prose?
7. ☐ Can a machine resolve who the brand is, where the product is made, and what claims are supported?
8. ☐ Can it resolve shipping cost, destinations, and delivery time?
9. ☐ Can it resolve the return window, conditions, refund method, and exclusions?
10. ☐ If you re-ran the same audit tomorrow, would you get the same baseline?

Reading the results: failures in 1–5 are product interpretability. Failures in 6–7 are brand. Failures in 8–9 are transactional. A failure in 10 means the instrument is not deterministic and its numbers cannot be used for before/after comparison — fix that before anything else.

What to tell a client

When they ask "isn't this SEO?"

> *"SEO makes you findable by people. This makes you interpretable by machines. Related neighborhoods, different work. Your organic search usually improves as a byproduct — because resolved categories and correct structured data are also what search rewards — but that's a side effect, not the objective."*

When they ask "will this get me ranked in ChatGPT?"

> *"Nobody can promise that, and anyone who does is selling what they don't control. What this controls is the first gate: whether a machine can correctly interpret your offer at all. Without that, the other factors never get their turn."*

When they say "our product pages are excellent."

> *"They probably are — for people. The two are independent properties. Rich, well-written pages routinely measure lower on machine-facing readiness than thin ones, because the information is there and simply isn't resolvable. That's the gap we measure."*

When they ask "why is this field empty?"

> *See "The rule that protects you," above. Say it exactly.*

Six mistakes

1. **Marking up ambiguity.** Adding valid schema to an unresolved product decorates the gap. Resolve first, express second.
2. **Enriching before resolving.** You get eloquent descriptions of a product nobody can categorize.
3. **Filling gaps with plausible values.** It looks complete, it passes validators, and it is a liability wearing the client's name.
4. **Rewriting human copy for machines.** You traded conversion for legibility, and you didn't have to — they are different fields on the same product.
5. **Comparing scores across instrument versions.** The scale changed; the "lift" is arithmetic, not improvement.
6. **Selling a projection as a result.** The single fastest way to lose a client in month three.

The discipline is open

IEO is not trademarked and no vendor owns it. You can practice it manually, with enterprise systems, or with any tooling — including your own. If you build an instrument, a validator, a benchmark or a counterfactual tester, it is yours.

The discipline does not require anyone's tools. It requires the measurable objective they share.

IEO optimizes understanding.

Full definitions: nsolvia.com/ieo/glossary
The founding paper: DOI 10.5281/zenodo.22104281
Suggested citation: López Castaño, J. C. (2026). *IEO Field Guide.* NSOLVIA Research. Version 1.0.