



PHASE 4 OF 6:
BUILDING A MODERN TAM

The Data You Pulled Is Not the Data You Think You Have.

Why third-party vendor classifications are fiction and how AI segmentation turns a raw market universe into a defensible TAM.

The Classification Problem No One Tells You About

By the time you reach Phase 4, you've done a lot of work. You enriched your client data. You ran the pattern analysis. You built guardrails grounded in actual distribution, not whiteboard intuition. You pulled a broad market universe - 90,000 companies, maybe 140,000 - because you understood that filtering early is just pre-qualification dressed as list hygiene.

And then you look at the industry field in your export and realize the problem.

Apollo has a classification for "beauty." So does ZoomInfo. So does Crunchbase. They are not the same classification. A brand that lives in "beauty & personal care" in one platform shows up under "consumer goods" in another and "retail" in a third.

A B2B software company gets tagged as "technology" in one database and "information services" in the next. Nobody is wrong, exactly. They each built their own taxonomy for their own purposes, and those purposes were never "make this data useful for your ICP model."

This is category drift and it compounds in proportion to the size of your universe.

At 1,000 records it's a nuisance. At 90,000 records it's a structural problem. If you skip segmentation and go straight from market universe to scoring, you will apply the guardrails you spent Phase 3 building against a population that doesn't reflect your actual target market. You'll score companies that were never relevant. You'll miss companies that were. And the TAM you end up with won't be wrong in any obvious, documentable way. It'll just quietly underperform.

The fix isn't better data vendors. It's defining your own categories and making the AI apply them.

Most teams enrich and analyze only their closest relationships



**The data you pulled is not
the data you think you have.**

What Your Definitions Need to Actually Do

Most teams think about segmentation as a labeling exercise. The AI reads each company, picks the closest category and voua la - done. That's not what's actually happening, and misunderstanding it is why the first run almost always comes back wrong.

The AI doesn't have your business context and context is king. It has a broad model of the internet as it existed in 2024. When you hand it 90,000 companies and ask it to classify them, it does exactly that using its general understanding of what words like "SaaS" and "fintech" and "professional services" mean to the general internet, not what they mean to your team when you're deciding whether to pursue an account.

The definitions you write aren't just category descriptions. They are the only business context the model has access to. Which means they need to carry the full weight of your ICP knowledge.

For each division you serve, you need: what the division is and isn't, what types of companies fall within each industry under that division, a concrete definition of what qualifies, example brands you know are good fits, close calls that still belong (and why they still belong despite being edge cases), and close calls that don't belong (and why they don't, even though they look like they might).

The last two categories are the ones most teams skip because they feel optional, but they aren't.

Here's what happens without them: the AI sees a company that shares surface-level characteristics with a category (similar language on their website, similar employee count, similar industry keywords) and classifies it there. Without explicit close-call guidance, the model has no mechanism for distinguishing a company that genuinely belongs from one that resembles the category without meeting the underlying criteria. Your false positive rate spikes and you get to spend manual review time cleaning up what the AI should have sorted correctly the first time.

The close calls are the definition of where your category actually ends.

What the AI Is Actually Doing During Classification

When you feed the segmentation prompt (your division definitions, industry definitions, example brands, and close-call guidance) and run it against your market universe, what you're running is a multiclass classification. **Each company in the list gets evaluated against every category you've defined and assigned to one, or flagged as outside the universe entirely.**

This is where the **freedom** of the model matters more than you'd expect.

The first version of the prompt you write will give the AI significant latitude. You've described the categories and given examples. But the model has room to interpret, and early runs tend to reveal where that latitude produces outcomes you didn't intend. Companies get placed in categories that technically contain the right keywords but are substantively wrong. Or the model starts creating sub-classifications you never defined, not because the AI is malfunctioning, but because it's filling in gaps in the guidance with its own logic.

The calibration cycle looks like this: **run the prompt, spot-check 15 to 20 results per category against your own judgment**, identify where the classifications diverge from what you'd have decided manually, and trace each divergence back to a gap in the prompt. Usually the gap is a missing close-call example or a definition that was accurate but not specific enough to exclude a look-alike category.

Then tighten the prompt and run again.

Each iteration is essentially removing latitude from the model not by restricting what it can see, but by giving it more context so it has less room to fill in gaps itself. The goal is a model that has enough business context to behave the way a well-briefed analyst would behave on the same task.

In practice, this **takes two to four rounds before the results** are reliable enough to scale.

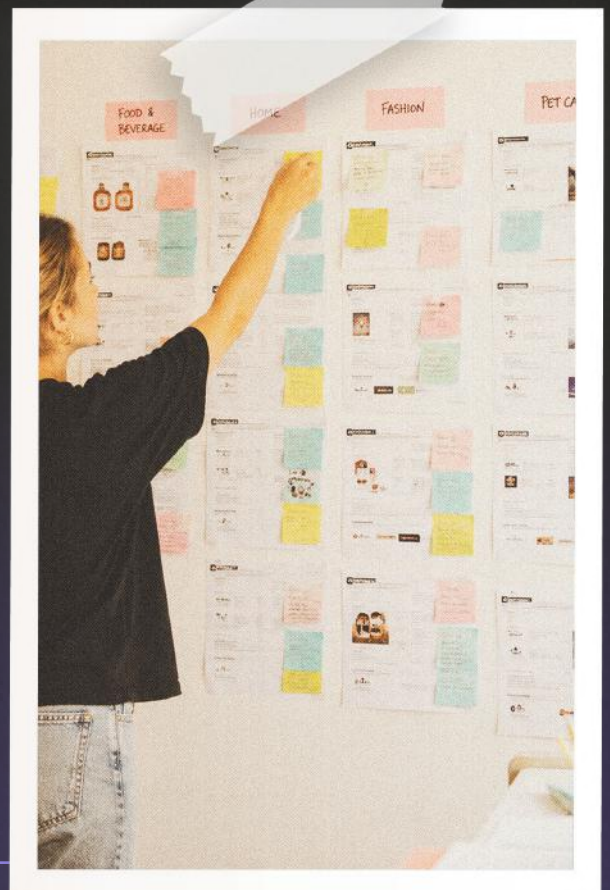
Why Manual Review Is Not Optional (and What It's Actually Reviewing For)

There's a tendency to treat the manual review phase as a quality check - a way of catching the AI's mistakes before they propagate. **That framing undersells what the review is actually doing.**

Yes, you're auditing for classification errors but really you're auditing for prompt gaps. Every misclassification is data about where your definitions were ambiguous, where your close-call examples didn't cover the actual edge case, or where a category boundary you thought was clear turns out to be genuinely contested. That information needs to go back into the prompt before the next run.

Manual review with prompt refinement is calibration and that's what makes the model scalable. Once the error rate on your spot-check drops below a threshold you can defend (most teams land somewhere between 5 and 10 percent), the model is ready to run at full scale.

The practical implication: **plan for this phase to take longer than you expect the first time, and shorter every time after that.** The prompt you build through calibration is an asset. The next time you run this process for a new market, a new vertical, an expanded ICP, you're starting from a calibrated template, not from scratch.



What Segmentation Unlocks

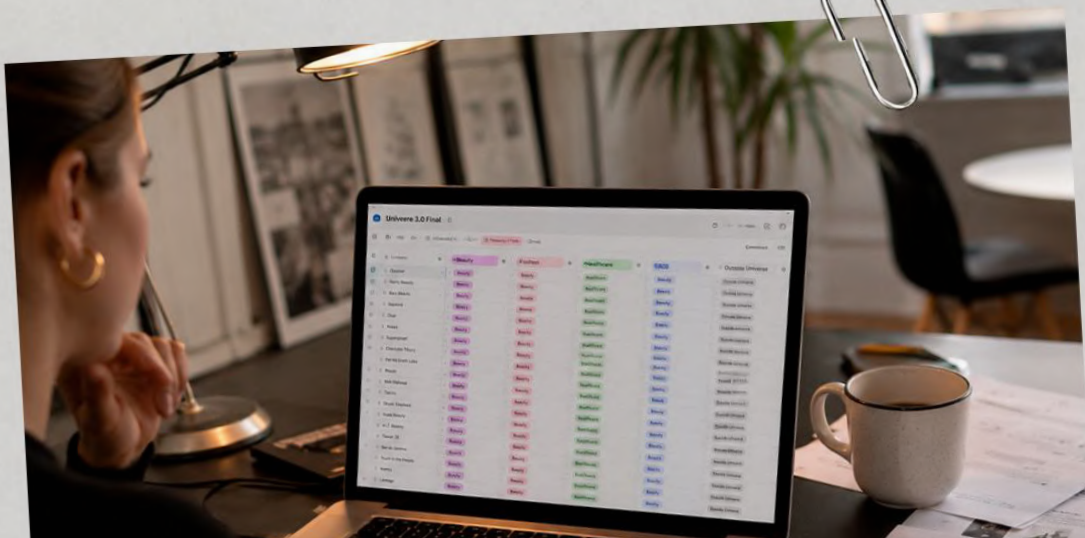
Once segmentation is complete, you have something you didn't have at the end of **Phase 3: a market universe that's been classified against your definitions**, not a vendor's. Every company in the list now has **a category assignment that reflects how your team actually thinks about the market**.

This is what makes scoring possible. The guardrails you built in **Phase 2 - the revenue floors, the headcount ranges, the vertical-specific signals** - were built from client data that already carried your category logic. When you apply those guardrails in **Phase 5, you're applying them to a universe that speaks the same language**. The thresholds mean what they're supposed to mean because the population they're being applied to is defined the way you defined it.

With that alignment your scoring produces a TAM you can defend not just to your team, but to the CFO asking why you're running media against this specific set of 14,000 companies and not a different set.

That answer exists in the documentation trail from Phase 2 through Phase 4. The guardrails came from the distribution analysis. The categories came from your definitions. The universe came from a pull broad enough to give the scoring rubric enough to work with. Every exclusion is traceable. **Every inclusion has a reason.**

That's the whole point.



Before You Move to Phase 5

Before you run scoring, verify three things: every company in your market universe has a category assignment or an explicit "outside universe" **flag; your prompt has been iterated to the point where spot-check error rates are defensible**; and your category definitions are documented (the full version, with close calls, not just the label names).

The scoring rubric in Phase 5 is only as precise as the segmentation feeding it. A well-calibrated segmentation handed to a scoring model produces a TAM. A poorly calibrated segmentation handed to a scoring model produces a very large, expensive list of companies someone will eventually have to explain.

This is Part 4 of a 6-part series on building a modern, data-driven Total Addressable Market.

Next: Phase 5 – Scoring, Qualification, and Building the Final TAM