

Is Your Synthetic Evidence Decision-Ready?

A practical checklist for marketing research teams

Plausibility is not validation. Synthetic respondents, digital twins, and AI-generated simulations can produce outputs that look convincing without necessarily being reliable enough for the decision at hand.

But that does not mean every synthetic use requires extensive validation. The right question is:

How much confidence do we need for this use, and what evidence will give us that confidence?

Use these five questions to determine what validation is warranted and what role synthetic evidence is ready to play.

1

PURPOSE

What are we asking synthetic evidence to do?

Start with the use, not the method. The amount of confidence you need depends on the role the synthetic evidence will play.

ASK

- **Are we exploring**, such as generating hypotheses, pretesting research, exploring scenarios, or iterating early concepts?
- **Are we prioritizing**, such as screening concepts, informing directional conjoint, modeling scenarios, or deciding where to invest additional research?
- **Are we informing a consequential decision**, such as a major launch, forecast, segmentation investment, or decision involving substantial financial or brand risk?
- Is synthetic evidence one input among several, or are we asking it to carry most of the evidentiary weight?

Why it matters: Synthetic evidence that is perfectly appropriate for exploration may not be sufficient for a consequential decision. Define its job before deciding how much validation it needs.

2

STAKES

How much confidence does this use require?

Once you know the purpose, determine the validation burden. It should rise with the consequences of getting the answer wrong.

ASK

- What happens if the result is wrong?
- How consequential is the decision it will influence?
- How reversible is that decision?
- Are we dealing with a population, behavior, or market condition where uncertainty or poor representation creates additional risk?
- What other evidence will be available before the final decision is made?

RULE OF THUMB

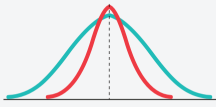
EXPLORE	LIGHT	PRIORITIZE	MODERATE	DECIDE	HIGH
For exploration, targeted checks may be enough.		For prioritization, synthetic evidence generally needs an appropriate anchor and robustness testing.		For consequential decisions, multiple forms of validation and human or real-world evidence should be expected.	

3

RISK

How could the synthetic evidence mislead us?

Do not simply ask whether the result looks reasonable. Ask how it could fail in a way that matters to your decision. Stress-test three recurring risks:

Bias concentration	Variability loss	Input + model fragility
Is the system systematically amplifying, suppressing, or normalizing particular tendencies? Look for patterns such as excessive optimism, novelty preferences, exaggerated group differences, or concentration around dominant brands.	 — Human spread — Synthetic spread Does the output preserve meaningful disagreement and heterogeneity? Synthetic systems can reproduce plausible averages while compressing variation and the tails of the distribution.	Does the substantive conclusion change when you make reasonable changes to wording, persona construction, context, model, version, or other configuration choices?

The test: Don't try to eliminate every difference. Focus on failures that could materially change the conclusion or decision.

4

EVIDENCE

What would give us confidence that the result is reliable enough?

There is no single validation test that works for every synthetic use case. In particular, simply demonstrating that an output resembles human responses may not be enough. Depending on the decision, look for one or more forms of evidence:

1 Human benchmark

Does the synthetic result reproduce the aspects of human behavior that matter, such as means, rank order, variance, segment differences, choice shares, or historical response patterns?

2 Structural or theoretical benchmark

Does the system behave consistently with established behavioral, economic, or marketing theory and known relationships?

3 Robustness

Does the substantive finding persist when prompts, personas, context, models, versions, or other reasonable conditions change?

4 Decision or outcome benchmark

Where a downstream outcome exists, does the synthetic result track actual behavior, market response, sales, adoption, or another consequential measure?

The principle: Benchmark the property you need to trust. Don't settle for evidence that the output merely looks or sounds human.

5

DECISION

What role has the evidence earned?

Now return to the purpose you started with. Based on the stakes, stress tests, and available evidence, decide what you are prepared to let the synthetic result do.

LEAD

Generate, probe, iterate

The use is exploratory, and the evidence is sufficient to move the work forward without additional validation.

SUPPORT

Screen, rank, allocate

The evidence is strong enough to influence prioritization or determine what happens next, with an appropriate human, behavioral, theoretical, or other anchor.

INFORM

Validate before the final call

The synthetic result can contribute to a consequential decision, but human or real-world evidence is required before making the final call.

ESCALATE

And ask one final question: What did we learn that would require us to escalate to additional human or real-world validation before proceeding?

CONTACT US

Want to learn more about MSI or the ARF? Email membership@thearf.org

