Questions to Ask Before Treating Synthetic Respondents as Decision-Grade in Choice Modeling
(Printable Reference)

Information Boundary

- What information do the synthetic respondents introduce beyond the original human data?
- What evidence shows that this new information reflects the world, not a hallucination?

Heterogeneity Visibility

- How did introducing synthetic respondents change the covariance structure of the model?
- When variance changes, how were those changes examined, explained, and understood?

Analyst Dependence

- Were the evaluation tests specified before synthetic augmentation results were examined?
- Which choices led to large changes in results, and which had little or no effect?
- Can analyst effects be operationally negated, or do results depend on analyst prompting and intervention?

Stability Under Removal

- Are results evaluated by progressively removing real respondents from a baseline with sufficient sample size and task depth, or only by observing when data are insufficient?
- Can the degradation path be shown as an artifact, not described in narrative form?

Extrapolation Discipline

- Are model conclusions extended to levels that were not observed in the human data?
- If so, what out-of-sample diagnostics demonstrate that these extensions reflect genuine generalization?

Failure Awareness

- If this approach created new risk, where would failure first appear first?
- Who is responsible for noticing, documenting, and communicating those failures?

Why These Questions Exist

If synthetic respondents are to become part of standard analytic practice, the standards governing belief advancement must mature alongside the methods themselves.

These digital twins are increasingly proposed as inputs to choice models supporting pricing, assortment, and design decisions. Their appeal is straightforward: speed, scale, and operational flexibility under constant analytic pressure. The central issue is not whether these methods can produce answers, but whether existing evidence standards are sufficient to evaluate them.

The risk is not technical failure, but confidence advancing faster than evidence.

Decision-Grade Requirements for Synthetic Respondents

The following requirements govern whether synthetic respondents may be treated as decision-grade inputs to choice models.

They sit above any specific modeling approach, validation procedure, or implementation detail, defining the conditions under which permitted declarable.

1. Artifact Rule

If you can't show it, it doesn't count.

Claims about synthetic respondents must be supported by inspectable artifacts: such as plots, tables, or pre-specified tests, rather than narrative assurances, procedural descriptions, reputation, or venue of presentation.

If it cannot be shown, it is an assertion, not evidence.

2. Baseline Integrity Rule

You don't test synthetic data on weak baselines.

Synthetic respondents may only be evaluated relative to a baseline that is adequate for decision use.

Evaluation must begin from a model supported by sufficient real data to inform the decision. Synthetic augmentation may be assessed by observing how conclusions behave as real data are removed from this baseline, not solely by observing improvements when data are already inadequate.

This rule distinguishes augmentation from substitution and prevents synthetic data from being used to manufacture apparent validity in weak designs.

3. Pre-Commitment Rule

No fishing, no quiet retries.

Evaluation criteria must be specified before synthetic augmentation results are examined.

Tests, thresholds, and comparison standards should be defined in advance, and unsuccessful tests may not be quietly discarded or replaced post hoc. This rule protects against result-conditioned evaluation and selective reporting while allowing legitimate probing and challenge by skeptical reviewers.

4. Invariance Rule

If it changes a strong answer, it's wrong.

If synthetic respondents materially change conclusions in a well-powered, decision-adequate baseline, the method is not decision-grade.

Synthetic augmentation may stabilize or interpolate where evidence is thin. However, when applied to a baseline that is already adequate for decision use, conclusions should remain invariant. Material shifts in optima, rankings, or recommendations under these conditions are disqualifying.

If any of these requirements cannot be satisfied, the method may still be promising. It should not be treated as decision-grade input to consequential decisions.

Appendix – Evidence Advancement Conditions

The questions on the first page define the boundaries that govern decision-grade use.

Those boundaries ladder up to three higher-order constraints that determine whether synthetic respondents are ready for use. Across all of them, assertions are insufficient: claims must be supported by artifacts rather than narrative reassurance.

These requirements do not prescribe methods; they govern evidence as analysis moves into action.

1. Evidence Boundary

Protects against treating synthetic augmentation as new empirical evidence.

This constraint governs whether synthetic respondents expand what is believed about the world.

- **Information Boundary** ensures that synthetic respondents are not implicitly treated as introducing new information without justification.
- **Extrapolation Discipline** requires that any generalization beyond observed choice contexts be explicit, theory-constrained, or independently validated. Each extrapolation must be justified on its own merits; prior successes do not transfer credibility.

2. Structural Visibility

Protects against invisible structural change being interpreted as improvement.

This constraint governs whether changes introduced by synthetic generation are visible before they are interpreted.

- **Heterogeneity Visibility** requires explicit inspection of how variance and covariance structures change when synthetic respondents are introduced.
- **Analyst Dependence** surfaces sensitivity to analyst choices, prompting parameters, and modeling decisions that may quietly shape results.

3. Decision-Grade Robustness

Protects against fragile conclusions being treated as stable guidance.

This constraint governs whether conclusions survive contact with data removal and failure.

- **Stability Under Removal** distinguishes signal supported by data from effects created by data expansion.
- **Failure Awareness** assigns responsibility for anticipating, detecting, and documenting breakdowns when assumptions fail.

These requirements do not assert that synthetic respondents are unsafe or invalid.

They require that methods claiming to fill data gaps demonstrate they are not hallucinating structure under sparse designs, limited task depth, or unobserved attributes or levels. Readiness for decision use depends on preventing confidence from advancing faster than evidence.

Hallucinated structure shifts apparent optima, causing decision error to compound as commitment increases.