PRINCIPLES TO FOLLOW FOR SUCCESS WITH SYNTHETIC SAMPLE

Neil Dixit
Founder and CEO, Glimpse
Email: Neil@glimpseahead.ai | Website: glimpseahead.ai
LinkedIn: www.linkedin.com/in/neildixit/

It's difficult to escape the often contentious debates about the value of "synthetic sample" or "digital twin" capabilities in market research these days.

Broadly speaking, synthetic sample relies on generative AI to create realistic responses to questions, based on the training data available to Large Models (LLMs).

In theory there are important benefits associated with the use of synthetic sample:

- Cost
- Time and efficiency
- Sample consistency
- Access to hard-to-reach or expensive audiences
- Privacy protections
- Real-time interactivity

Yet lots of people are skeptical about the wholesale replacement of human respondents with synthetic sample.

So am I!

You might reasonably ask, then, why the company I founded, Glimpse, is investing so much time and money in our synthetic capabilities (or what we call "Enriched Data").

It's because we believe that the debate about synthetic sample is based on a false choice between more traditional research techniques, on one hand, and newer gen AI-enabled approaches, on the other hand.

We see a third path emerging instead: using reliable synthetic sample (or Enriched Data) to extend and scale the value of human research and insights.

In practice, we think that the most successful approaches to synthetic sample across the industry are already starting to follow the same basic set of principles:

1. Building AI-generated data on a foundational layer of traditional, real-world, first-party data, including inputs like behavioral and demographic data and lots of rich human language data. (In a much-cited Kantar blog about synthetic sample, it's noteworthy that the experiment did not draw on first-party data at all to generate synthetic sample.)
2. Leveraging all the proprietary datasets your organization has available to create synthetic sample--rather than relying exclusively on commercially-available LLMs. (This is how organizations will use synthetic sample to achieve insights unavailable to competitors.)
3. Using synthetic sample to supplement (rather than replace) human sample, especially when it comes to hard-to-reach groups of respondents.
4. Aligning the creation and use of synthetic sample with specific marketing, innovation, content, and research goals. For instance, synthetic sample might be a valuable supplement to traditional sample when it comes to exploring broadly-held beliefs about brands but less reliable when it comes to eliciting reactions to truly innovative products.
5. Seeing the creation of synthetic sample as an opportunity to address bias by focusing on the representation of previously excluded groups of people in the foundational data.
6. Testing, testing, testing and embracing controlled experimentation.

Regardless, one thing is clear: You may not be interested in synthetic sample but synthetic sample is interested in you!

Increasing numbers of research firms and brands will incorporate synthetic sample into their market research toolboxes over the next year.

I recommend refusing to play the "pro/anti synthetic sample" game and instead establishing durable, foundational principles to guide the path forward.