The AI Tools That Are Transforming Market Research

Eight ways to experiment with "synthetic personas" and "digital twins."
by Jeremy Korst, Stefano Puntoni, and Olivier Toubia

HBR Staff/Tara Moore/Getty Images

Custom market research has always been notoriously slow and costly to conduct, often requiring many months and significant investments. As a result, marketers have made many strategic decisions without the benefit of timely external insights. But as we argued in our recent *HBR* article**, "**How Gen AI is Transforming Market Research," gen AI is transforming the collection, creation, and analysis of consumer and market insights—the lifeblood of strategic marketing. We're not alone

in making that argument: Since our article's publication, for example, both Andreesen Horowitz and Foundation Capital have published investment theses predicting that gen AI will dramatically transform the $140 billion global market-research industry.

At the center of this transformation, we believe, will be two AI-generated tools designed to act as proxies for people: the *synthetic persona* (representing a composite individual or group) and the *digital twin* (representing a real individual). By using publicly available or proprietary data to simulate human responses to questions and surveys, these new tools promise to allow marketers to conduct research and experiments without the time, cost, and participant burden of traditional interviews or surveys. Both thus augment what we call the "oracle" approach to gen AI use in marketing, in which researchers simply ask a standard LLM a raw question of interest (e.g., "What is the size of the addressable market for this product?") and hope to obtain an answer that can inform decision-making.

In this article, we'll discuss the nature of each tool, provide some examples of how they're being used by companies today, and then offer some step-by-step guidance to leaders on how to navigate this complex and exciting space.

The Synthetic Persona

In this approach, researchers, marketers, product managers, and even sellers provide an AI model with demographic, psychographic, or behavioral information about a customer type or segment to create a persona that's representative of that type or segment. They can then ask the LLM to answer questions as if it were that type of person. Supporters of this approach often note that many if not most strategic product and marketing decisions are based on a segment's combined insight rather than individuals.

Researchers can elicit responses from such synthetic personas in multiple ways. In one, which we call the top-down approach, they ask the composite persona to generate a point estimate or range for a segment—for example, by asking, "What would be this segment's average willingness to pay for the product?" This is a sort of AI super-agent, one that you can rely on always to give you a single "best" answer. Alternatively, in what we call the bottom-up approach, they elicit a range of responses by creating a whole population of simulated consumer personas who match the segmentation criteria. This group of synthetic personas is sometimes referred to as a "silicon sample." These are not super-agents optimized to give you the "best" answer for the segment; instead, they're designed to intentionally vary from individual to individual, just as the people within any given market segment do. Researchers can then ask these populations questions and aggregate their answers, just as they do in traditional market research.

Based on our observations and work with our clients, many researchers are often more comfortable with the bottom-up approach than with the top-down approach. Why? Because there's variability in the responses, and in general the process seems closer to how traditional market research works. (Recent academic research shows that people tend to prefer algorithms that make predictions in a way that matches the process they are meant to simulate.) However, it's still an open question whether one approach or the other works better, and for what.

For researchers who are interested in modeling differences across consumers, why not try to replicate the unique behavior of actual individuals to fully capture heterogeneity in the market? This is what digital twins, discussed next, are all about.

The Digital Twin

This approach goes beyond composite segments and imagined personas and brings individual-level data to bear on the task. If a company has detailed, individual-level data about a group of customers from past interactions, surveys, or interviews, for example, it might use that data to create digital twins of these individuals, who can then take part in virtual surveys, qualitative "interviews," and even experiments in the place of the individuals they represent. This has many interesting potential applications—and some serious ethical implications. Given the exploration going on in this area, it's possible that somebody may *already* have created a basic digital twin of you and is using it to optimize and personalize sales materials or a marketing message.

Some of the most ambitious work in the space of digital twins is happening at Columbia Business School, where one of us (Olivier) is working with colleagues on the Digital Twins Initiative. The DTI team, whose work is all open sourced, has to date created more than 2,000 digital twins of real people by having them respond to four waves of surveys with more than 500 questions covering a combination of standard demographic, psychological, behavioral economic, personality, and cognitive measures. After evaluating a variety of approaches to constructing the twins with the underlying data, the CBS team found that their best twins had a relative accuracy of 88% compared to their human counterparts in a test-retest benchmark —a very promising result. At the same time, digital twins replicated only about half of the experimental effects found among humans when replicating 17 well-known behavioral economics experiments, suggesting that further testing was warranted.

With this initial foundation, the Columbia team decided to dig further, conducting 19 additional studies across a range of domains, testing both the individuals and their digital twins. These studies were selected to be

representative of the use cases in which digital twins might be deployed today. Then the team conducted a meta-analysis, which identified areas of potential for digital twins—notably, questions relevant to social interactions (e.g., fairness of various advertising or pricing decisions, capturing the impact of personality traits on purchase decisions) or human-technology interactions (e.g., consumer behavior on technology platforms).

The meta-analysis also identified some important limitations. For example, digital twins were less able to capture the diversity of human opinions in political domains, and they were more likely to provide socially desirable answers. Digital twins also tended to display an interesting mix of "pro-human" bias (e.g., more likely than their human counterparts to believe that people are fair and can be trusted) and "pro-technology" bias (e.g., more accepting of algorithmic hiring and more tolerant of online targeting). And they were more accurate, by a couple percentage points, when predicting responses of more-educated, higher-income, and ideologically moderate participants.

How did the performance of digital twins compare to that of synthetic personas? The researchers compared their digital twins to synthetic personas, based on an extensive set of demographic characteristics, across the 19 studies. Digital twins were better able to capture variations across participants, as measured by the correlation between the true and simulated answers. While the correlation was modest (0.2 on average), additional analyses suggest that it might not be realistic to expect a correlation higher than 0.3 in this context. When it came to predicting the exact answers given by participants (as measured by accuracy), digital twins and synthetic personas were tied at 75%. This suggests that although the rich individual information contained in digital twins helps capture variations across people, it does not necessarily help predict each single answer more accurately.

Ultimately, the CBS team concluded that "while digital twins show promise, they are not fully 'ready for prime time' yet."

This discussion hints at an issue that often does not receive sufficient attention: It may be not be obvious how to judge the quality of synthetic data. If managers have access to individual-level data, they can estimate accuracy by measuring how close each synthetic response is to its human counterpart. Alternatively, they can assess correlation—that is, how well variations among synthetic responses reflect variations among human responses. If managers lack access to individual-level data to match each synthetic response to its human counterpart, they can estimate the accuracy of the average synthetic response or compare of the overall variation in synthetic vs. human responses. Some managers may be more interested in knowing whether relying on synthetic data would lead to the same versus different decisions, but we currently don't have well-established metrics for such a benchmark.

How to Get Started

Academia and industry are just beginning to learn how to leverage gen AI for strategic insights. This area is rapidly advancing, but it's likely to be some time before we have a solid grasp of what works for what purpose and why, so for the near future we expect traditional research to remain the primary method for collecting insights.

That said, now is the time for companies to start experimenting and testing possible use cases to understand how LLMs and synthetic data can augment and enhance their decision-making. Here's how we recommend you go about that.

1. Determine your use case. Keeping the end in mind is crucial: Are you seeking a point estimate from a key segment (e.g., likelihood to buy), or do you want to understand how individuals differentially value certain

product features? Are you trying to augment the results of a traditional survey with hard-to-reach individuals? Do you need quantitative results only, or are you looking for qualitative feedback?

2. Determine the type of consumers you want to simulate. Do you want to better understand the average or composite point of view of a particular segment (e.g., manufacturing finance director, college-aged male), where a synthetic persona might work? Or do you want "feedback" from a specific consumer or set of consumers (e.g., 500 midwestern mothers between the ages of 24 and 36), where it may be worth the extra effort and cost to create digital twins that can capture variations and nuances both within and across segments.

3. Get the data necessary for calibrating the personas or twins. This can be the limiting factor for some options. If you have a trove of past data about specific consumers, then creating digital twins becomes a more viable option. However, if you are limited to aggregate or publicly available data, then synthetic personas might be a more realistic option. In either case, you will need to experiment with your twins or personas to evaluate just how much data is needed, and how much may be too much.

4. Decide on clear and precise performance metrics. Although it is tempting to treat performance metrics as interchangeable, they are not, and the choice of your performance metric will influence the conclusion you will draw from your data. Hence, to avoid cherry-picking after the fact, we urge researchers to commit to a set of metrics before collecting data.

5. Run a small test. Ideally, this test should be representative of your use case, and you should run it on a sample of human participants who

are representative of your customers. This data provides a measure of "ground truth" against which you will compare synthetic data.

6. Evaluate the performance of your synthetic personas or digital twins. Compare your synthetic data to that measure of ground truth. Make sure to include relevant benchmarks that will help you contextualize the results. For example, how well does a random benchmark perform? How about other simpler benchmarks that use less individual-level data? An absolute measure of performance (e.g., 80% accuracy) is not as useful as a comparison between relevant benchmarks, which allows you to explore the tradeoff between cost/complexity and performance.

7. Decide whether you want to scale. By this stage of the process, you will have a holistic view of the potential of synthetic data for your use case. A cost-benefit analysis will allow you to decide whether it is worthwhile to include synthetic data in your research pipeline.

8. Periodically test to ensure ongoing validity and keep your data fresh. For example, it may be useful to periodically run parallel studies on the real humans behind your digital-twins panel to ensure continued performance and to augment your training data with up-to-date human data.

. . .

Given the growing interest in this space, we will undoubtedly see increased interest in synthetic personas and digital twins. They may not be quite ready for prime time yet—but they will be soon, and when they are they're likely to revolutionize the world of marketing. Smart companies should therefore start experimenting with and investing in them now.

This article was originally published online on November 17, 2025.

Jeremy Korst is the founder of Mindspan Labs, an AI-transformation consultancy and incubator. He previously held senior executive roles at Microsoft and T-Mobile. He also serves as partner at GBK Collective, which provides strategic insights to global brands. He is a co-author of the Wharton School's annual study on enterprise AI adoption.

Stefano Puntoni is a professor of marketing at the University of Pennsylvania's Wharton School and is the co-director of Wharton Human-AI Research. He is one of the leading behavioral scientists studying the business applications of AI.

Olivier Toubia is a professor of marketing at Columbia Business School and a recognized leader in the field of quantitative marketing.