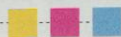




AI & MACHINE
LEARNING

How Gen AI Is Transforming Market Research



A guide to the
most promising
opportunities.



AUTHORS

**Jeremy
Korst**
Partner,
GBK
Collective

**Stefano
Puntoni**
Professor,
Wharton
School

**Olivier
Toubia**
Professor,
Columbia
Business School



ARTIST CARSON DAVIS BROWN







AI & MACHINE
LEARNING

Among all the managerial functions, marketing is likely to be the one that's most disrupted by generative AI.

Recognizing the technology's potential, academics and practitioners alike have been investigating new ways to apply it to customer service and content creation for some time now, but recently the business world has started paying attention to the impact it could have on other marketing activities. The most exciting of these is market research, the processes by which firms gather data and generate insights about customers and competitors.

For the past two years we've studied and worked directly with companies that are exploring the use of gen AI in market research, and we can report that big, transformative changes are in store in that arena. When properly deployed, the technology offers firms unprecedented opportunities to understand and engage with customers, better assess the competitive environment, and extend data-driven decision-making deep into their organizations.

In our research we've identified four distinct classes of opportunities. The first involves *supporting current practices* by making them faster, cheaper, or easier to scale up. The second involves *replacing current practices* by leveraging



ABOUT THE ART

Inspired by the tradition of American quilting and performance art, Carson Davis Brown assembles guerrilla installations of found objects in big box stores.

synthetic data (data about people's preferences or behavior that's created by AI and not gathered through surveys or interviews). The third involves *filling existing gaps in market understanding* by obtaining insights and evidence that aren't available in conventional data. And the fourth, which is still just emerging, involves *creating new types of data and insights*.

In this article we'll lay out a framework that will help guide leaders and companies through this new terrain. We'll explain how gen AI has started and will continue to change market research, what can be done to make the most of new opportunities, and how to be alert to the technology's limitations and to the ethical issues surrounding its new applications.

Supporting Current Practices

Firms often are frustrated by the relatively high cost and long timelines of collecting customer and market insights. So how might gen AI address both?

Drawing on what we've learned in our research and company engagements, we've developed a systematic way to find use cases. It involves applying what we consider to be the four core capabilities of gen AI—synthesis, coding (computer programming), human interaction, and writing—to each stage of the market research process. For simplicity's sake, we've reduced the process to just three stages: the identification of opportunities and design of a research program; the collection and analysis of data; and the reporting and dissemination of information and insights. (See the exhibit "How Gen AI Can Enhance Current Market Research Practices.") Gen AI's ability to synthesize information, for example, could be leveraged to summarize literature and previous research in the first stage, to extract findings from interviews and new data in the second stage, and articulate takeaways in the third stage. And gen AI could do all those activities far more rapidly than humans could.

How are firms actually leveraging gen AI to support their current practices? A survey that we recently conducted with Jeremy's company, GBK Collective, a marketing strategy and insights firm with deep ties to top business-school professors and *Fortune* 500 companies, reveals how. In it we gathered responses from more than 170 market research practitioners and users. We learned that 45% of them were



AI offers firms unprecedented opportunities to understand and engage with customers, better assess the competitive environment, and extend data-driven decision-making.



IDEA IN BRIEF

THE OPPORTUNITY

Generative AI is poised to transform market research—the way firms gather data and develop insights about customers and competitors. Marketers need to understand and take advantage of the possibilities.

THE RESEARCH

The authors have identified four distinct application classes for gen AI in market research: (1) supporting current practices, (2) replacing current practices, (3) filling existing gaps, and (4) creating new types of data and insights.

THE WAY FORWARD

As the marketing community engages with gen AI, it will need to embrace the power of the new technology while also ensuring that the strategies it devises are grounded in fairness, accuracy, and genuine customer insights.



AI & MACHINE
LEARNING



One of the most innovative applications of gen AI in marketing is producing and analyzing what's known as "synthetic data."

already employing gen AI in their current data and insights activities; another 45% told us they were planning to do so in the future.

The survey surfaced some interesting top-level trends. Not surprisingly, more than 70% of respondents reported concerns about the possible side effects and challenges of gen AI. Those concerns included the potential for biased or inaccurate information, security and privacy risks, and the additional time and energy it would require to integrate gen AI into current practices. A similar percentage of respondents had concerns about gen AI's potential to create skill gaps and even replace human data and insights professionals.

That said, many respondents—like many of the practitioners we've worked with—were overwhelmingly positive and already embracing the technology. Sixty-two percent of those currently employing gen AI in their work told us they were using it to synthesize lengthy interview transcripts and other documents, a formerly laborious process; 58% were using it to analyze data; and 54% were using it to write reports. On the whole, our respondents seemed excited about the many ways gen AI could help them get things done. More than 80% agreed that it has the potential to significantly enhance personal productivity and efficiency and that integrating it into their work processes is critical for staying competitive. An equal percentage told us they believed that it will positively affect their industry overall by improving their jobs and driving significant innovation. It will do so, they said, by enabling people to perform tasks faster, focus on tasks where they can add more value, and carve out more time for data interpretation and storytelling. More generally, it will increase the quality, accuracy, and customization of their work.

Market research startups are already moving into this space. One of them, Meaningful, for example, aims to "supercharge" market research by using it to create customized surveys, distribute them to panels of participants, conduct qualitative interviews, and analyze the results. Outset.ai focuses on the ways gen AI can generate questions rather than answers about consumer needs and behaviors—a promising idea. Its AI-moderated research platform dynamically probes participants with new questions that are based on their previous responses to get more-insightful answers from them, combining the speed and scale of an automated survey

tool with the depth of a traditional interview. Aaron Cannon, Outset.ai's cofounder and CEO, has built his company on a singularly important finding: When gen AI has the simple job of engaging people in conversation, they share their thoughts, experiences, and feelings abundantly, and AI's hallucination problem virtually disappears.

The research team at WeightWatchers (an Outset.ai customer) has found that participants also are often more forthcoming when being interviewed by AI rather than by people, because certain bias effects are reduced. Wil Readinger, the former head of user experience research at the company, is full of enthusiasm for gen AI. No longer, he says, will researchers be forced to choose between "the richer, contextualized data collected in an interview and the broader reach of a survey." Instead, they now have a third option, which, as he puts it, is "Both!"

Replacing Current Practices

One of the most innovative applications of gen AI in marketing is producing and analyzing what's known as "synthetic data"—artificially generated data that mimics real people's behaviors and preferences. Firms can do this with any of the widely available gen AI programs, but they can also develop and train their own specialized models using the aggregate data that they've already collected from traditional research, syndicated data, CRM systems, and transactional information. The synthetic data can then be used to simulate various customer or competitor responses, highlighting potential pain points and the benefits consumers seek at different stages of their interactions with a product or service. A full 81% of the respondents in our survey told us they already use or plan to use gen AI to create synthetic data. One mentioned plans to "create synthetic audience personas my team can interact with that match my client's target audience."

By creating detailed hypothetical customer profiles and scenarios, these models help marketers anticipate needs and preferences better and make more-effective decisions. But they certainly can't replicate the full depth and unpredictability of human behavior. Our respondents were very aware of that shortcoming: Only 31% rated the value of data produced by gen AI as "great"—making it one of the lowest areas of satisfaction in our survey. Significantly, though,

How Gen AI Can Enhance Current Market Research Practices

A simple tool for identifying use cases

	OPPORTUNITY IDENTIFICATION & RESEARCH DESIGN	DATA COLLECTION & ANALYSIS	REPORTING & DISSEMINATION
SYNTHESIS	Summarizing existing research and literature	Extracting meaning and insights from text	Articulating takeaways
CODING	Mining existing data for hypotheses	Setting up surveys in a web interface; performing analytics	Creating data visualization tools
INTERACTION	Using chatbots for brainstorming and idea generation	Using synthetic interviewers to ask follow-up questions	Using chatbots for data exploration
WRITING	Generating hypotheses	Creating study materials	Crafting executive summaries

studies have shown that gen AI can improve the quality of synthetic data when it's fed question-and-answer examples in prompts, when it's allowed to leverage a knowledge base from past studies to retrieve content relevant to each question, and when its parameters are fine-tuned to better fit the existing data.

New and established companies alike are exploring the potential of synthetic data. The startup Evidenza, for example, has conducted more than 60 validation studies comparing synthetic results against traditional research across multiple industries. In one study it collaborated with EY to conduct a double-blind test—the gold standard in scientific research—in which neither party knew the other's results until the study concluded. EY provided Evidenza with its annual brand-survey questionnaire and details about its target audience (CEOs of U.S. companies with more than \$1 billion in revenue) but withheld the actual survey results for later use as a benchmark. Evidenza then created more than a thousand synthetic personas that matched the target-audience profile and had them answer the survey. “The results were astonishing,” Toni Clayton-Hine, the CMO of EY Americas, told us. “The conclusions were 95% the same, the correlations were very strong, and in many cases the numbers were nearly identical.”

Gen AI doesn't produce only structured quantitative data. A 2024 study done at the Wisconsin School of Business demonstrated that it can yield deep qualitative data too. In particular, it can conduct revealing, insightful interviews

of the synthetic respondents it has created—for example, to replicate desired customers. Several researchers and managers have already successfully used it that way. A good example of a vendor that offers this sort of work is the Portugal-based startup Synthetic Users.


Of course, any company that wants to pursue a tailored approach to creating synthetic data has to share some proprietary information with the gen AI program it uses, and that makes some firms uncomfortable. To ease their concerns, all the major gen AI providers offer paid enterprise versions of models that won't share proprietary company data or insights with other firms. Some are also helping companies create “small” gen AI models that they themselves will entirely control. The startup Rockfish Data, for example, allows firms to develop and train their own custom gen AI models on in-house datasets—an approach that keeps data and models completely private. These models may be small, but some of the users are not: The U.S. Army and the Department of Homeland Security are both Rockfish Data customers.

Again, there are positives and negatives to think about: Small gen AI models are mostly limited to structured or semistructured data (data that's numerical or categorical) and don't benefit from the public models' broad training sets, whereas public models can also work with less-structured qualitative data. For some companies fine-tuning large models using proprietary data can be an effective compromise.





AI & MACHINE
LEARNING

 Gen AI can conduct insightful interviews of the synthetic respondents it has created—for example, to replicate desired customers.

Filling Existing Gaps

Even in organizations that profess to be data-driven, practitioners often report that most decisions are made without a formal empirical analysis. There's simply not enough time or money to do one. But gen AI promises to be an always-on intelligent engine for customer and market insights—one that can offer market researchers instant access to empirical evidence when data isn't available or is too costly to acquire. Gen AI can be used to test assumptions, pilot concepts and execution strategies, and provide a sounding board for managerial decisions. Firms can even develop “labs” that make customized AI models available to employees in a safe and convenient way to support decision-making throughout the organization.

In our survey 30% of respondents said that their company had used gen AI to guide decision-making that previously wouldn't have leveraged external data and insights. Overall, 81% of respondents reported using or planning to use gen AI to “listen to the market” and keep their organizations informed about the competitive environment. One, for example, used it to analyze the latest trends and rivals' strategies and produce timely competitive intelligence for decision-making, and another used it to perform predictive analytics for decisions, drawing on historical data and assumptions.

Many companies are experimenting with using synthetic data to support product innovation. One of them is General Mills. “We're exploring how synthetic data could accelerate and improve our product-ideation processes, increasing the likelihood of finding truly great ideas about how to best serve our consumers,” says Lanette Shaffer Werner, the company's chief innovation, technology, and quality officer.

A number of startups are getting involved in synthetic data too. Evidenza is providing tools to create it on B2B customers, who are notoriously hard to reach. Arena Technologies is using gen AI and synthetic data about local customer profiles and tastes to help retailers make smarter decisions—about how to tailor offerings by outlet, for example. Evidenza employs synthetic data to help marketers make decisions about targeting, positioning, and messaging, and its platform also estimates the financial impact of those choices, with ROI projections and metrics that speak to CFOs and revenue teams.

Creating New Kinds of Data and Insights

A mantra in content marketing and sales is that you get only one chance to make a first impression. But maybe that's no longer true.

We say that because content marketers and salespeople are starting to use gen AI to create “digital twins”—virtual replicas of individual customers that are constructed using publicly available information or proprietary data—to test and refine their materials and pitches before presenting them to real people. This approach allows for the meticulous calibration of marketing efforts because digital twins, unlike actual people, never get tired, irritated, or bored when interacting with marketers and their questions. More than 40% of our respondents said that they're already experimenting with digital twins. One, for example, reported using digital twins in a virtual sales environment to “simulate customers' purchasing behavior, click-through rates, and interaction patterns in different contexts,” all in an effort to “help test-market strategies and optimize the user experience.” Another 42% said that they planned to experiment with digital twins in the future.

The use of digital twins in marketing is mushrooming. Arena has built a training tool that B2B sales reps can use to interact with digital twins of customers. CivicSync has developed a technology that allows its customers (with consumers' consent) to track shopping, search, and other online behaviors and then build highly precise digital twins of their target users. The PR firm Ogilvy has tried creative ideas out on digital twins to ensure that its campaigns will resonate with consumers. And GBK Collective is experimenting with different ways to use survey results to train or prompt gen AI to create digital twins that can be consulted on follow-up marketing questions. It uses different subsets of past survey data to create different variants of digital twins and then runs tests to see if some work better or worse for certain business research goals. The test results get compared with the actual responses from the past surveys to measure the effectiveness of each approach.

Many firms are also experimenting with freely available tools like Google's NotebookLM, which creates a personalized “research assistant” that is trained with information about competitors, relevant industry and domain data, and



AI & MACHINE
LEARNING



Bias is intrinsic to any training dataset and can skew AI outputs, leading to misrepresentations of customer segments or market trends.

profiles of target customers. The assistant can prepare team members for customer engagements by helping them refine their pitches, offers, and interactions and anticipate potential objections. Henry Sosa, a principal technical account manager at Oleria, a cybersecurity startup (where one of us, Jeremy, is an adviser), has already created a host of these gen AI assistants for his sales and marketing colleagues.

Academics, too, are turning their attention to new possibilities. A team at Columbia Business School, for example, is building a representative panel of 2,500 personas, each the digital twin of a real person. The people they will mirror will go through an extensive battery of tests (psychological, behavioral, cognitive, attitudinal) that will collectively establish a “ground truth” that gen AI can then use to create the twins. The idea is to deploy the panel as virtual subjects in new research and surveys. A study from a team that includes researchers at Stanford and Google DeepMind suggests that this approach has promise. After interviewing a sample of individuals for a couple of hours and having them complete a series of surveys, the team used the interview transcripts to create digital twins of each participant. Then it asked the digital twins to respond to the same surveys. The real participants were also asked to answer the same questions again two weeks later. The digital twins’ responses turned out to replicate the real people’s initial responses 85% as accurately as those people’s second set of responses did.

Understanding the Limitations

Gen AI offers marketers a lot, but it still has plenty of limitations, which are important to acknowledge. As we noted earlier, one of the main concerns that our survey highlighted was the potential for biased results, which was cited by 77% of respondents. Bias is intrinsic to any training dataset and can skew outputs, potentially leading to misrepresentations of customer segments or market trends. (However, current survey practices can also lead to biased results for a variety of reasons.) Further, given that gen AI models are trained on existing data and insights, it’s not yet clear how good they’ll be at predicting dramatic changes in consumer behavior or anticipating discontinuous product innovations. Gen AI models are also known to be sensitive to prompt

architecture. For example, we’ve shown that when they’re answering multiple-choice questions, they’re influenced by the order and labeling of the options in ways that can be unpredictable. Researchers should be aware of this effect and, as they do with real people, make sure to randomize all relevant aspects of the survey to limit potential bias.

Concerns have also been raised about gen AI’s ability to simulate responses from a representative sample of the population. One 2023 study by researchers at Columbia and Stanford found that most recent models from OpenAI express opinions that are more typical of people who are, for example, liberal or well-educated, and are less characteristic of people who are over the age of 65 or more religious. Such bias may come not only from the training data on which the models are built but also from human involvement in refining them, which would explain why more-recent models are proving to be increasingly biased. Given those limitations, it’s perhaps not surprising that a 2024 study led by James Bisbee of Vanderbilt University found that when synthetic respondents took a public opinion survey, their responses closely resembled humans’ answers but showed less variation, were sensitive to the wording of the questions, and were not stable over a three-month period.

Synthetic data can also be of limited help in the simulation of experiments that assign respondents to different treatments across conditions. We’ve explored the use of gen AI in simulations of experiments in which the price of the product varied across digital respondents, who were asked whether they intended to buy it. We’ve found that the demand curves elicited by the AI not only were different from the curves elicited by asking the same questions to human respondents but also were implausible.

The reason is perhaps simple in hindsight. Experiments that involve real people typically employ blind designs, where each participant is unaware of the other conditions and of the goal of the study. That is what happens when consumers make decisions in the real world, where they typically are not offered a choice of the exact same product at different prices. But when applied to experiments involving gen AI models, blind design leads to a phenomenon called “confounding,” whereby an unforeseen variable associated with both the treatment (say, the price) and the dependent variable (say, the decision to purchase) muddles the



relationship between the treatment and the dependent variable. In our pricing experiments we found that gen AI had a hard time treating the price of the product as random and tended to assume that other variables (past prices, prices of other products in the store) also varied with the price of the product, which justified purchasing at a higher price, leading to little change in behavior as price varied.


Gen AI also still doesn't predict consumers' emotional reactions to multisensory stimuli well. That's what the flavor and fragrance company Takasago learned when it explored replacing a human research panel with an AI-generated panel. Fragrances, for example, just don't have the same effects on gen AI models that they do on real people. "While our model can learn from the past data," we were told by Jelena Le Breton, the company's global vice president for customer insights and market research, "it is still hard to predict for human change of olfactive preferences, which can often be associated with individual experiences and emotions."

AS WE NAVIGATE the future of marketing in the era of gen AI and synthetic data, the marketing community will need to proceed with a balanced and informed perspective. By

doing so, we can harness the power of these remarkable tools and technologies while ensuring that our strategies are grounded in fairness, accuracy, and genuine customer insight.

With that in mind, we'd like to conclude on an optimistic note, by bringing up an idea that Cannon of Outset.ai proposed to us. If gen AI can talk to thousands of people around the world in hundreds of languages every hour, and it can instantly draw all sorts of unique, high-fidelity insights from the data those conversations generate, then our understanding of one another should deepen—and the products, services, and experiences we build will be infused with more humanity, not less. ■

HBR Reprint R2503H

 **JEREMY KORST** is a partner at GBK Collective, a strategic-insights, consumer-behavior, and analytics consultancy. He previously held senior executive roles at Microsoft, T-Mobile, and Avalara, among other companies. **STEFANO PUNTONI** is a professor of marketing at the University of Pennsylvania's Wharton School and is one of the leading behavioral scientists studying 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.