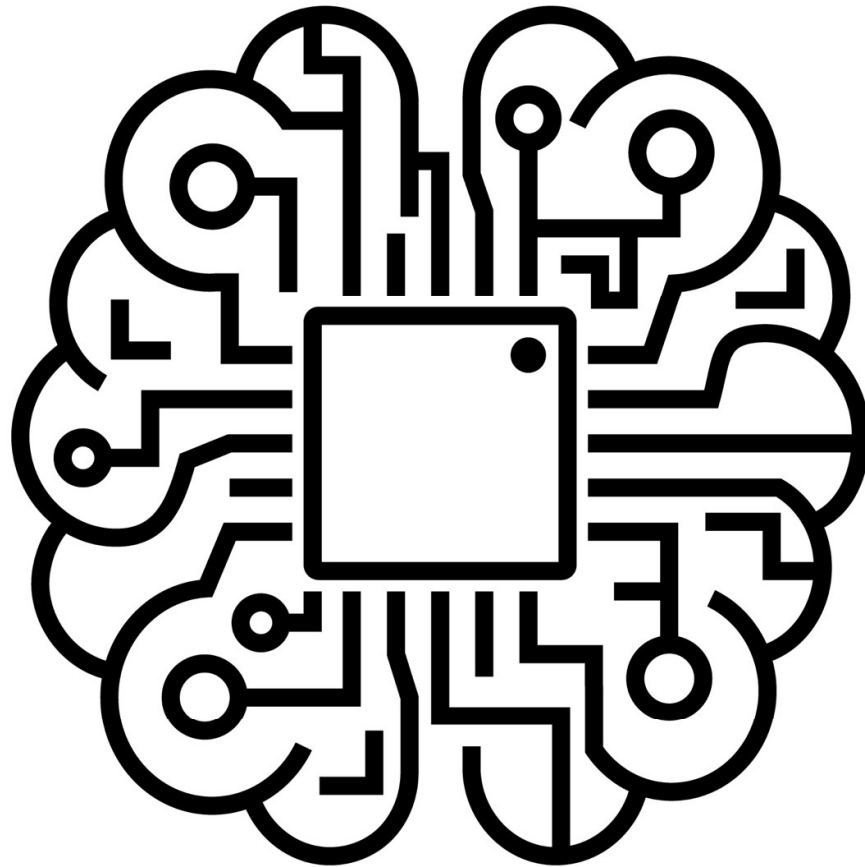




---

**Advanced Insights and Analytics Council**  
**AI and Synthetic Data**  
**in Market Research**

Authors:  
Sam Barton  
Jarrod Howard-Payne  
Anthony Shephard-Williams  
Jake Steadman



## Introduction

In the Generative AI boom, one area gaining notable attention is the creation of “Synthetic Data” in market and marketing research. An [article by Mark Ritson](#) has recently boosted interest in this area, but it has been a topic of conversation for several years. This paper explores how AI and synthetic data can be utilised in the market research industry, examines how this change can be leveraged for competitive advantage and provides a framework for further exploration.

## Synthetic Data

AI is the hottest tech in the world; [80% of Fortune 500](#) companies mention AI in their earnings reports. There are several ways that AI will influence the market research industry, but one of the most fundamental is the creation of Synthetic data.

“Synthetic data is artificially generated information designed to mimic real-world data. It is created using sophisticated techniques to replicate real data's statistical properties and structure, but without containing any actual real-world observations.”

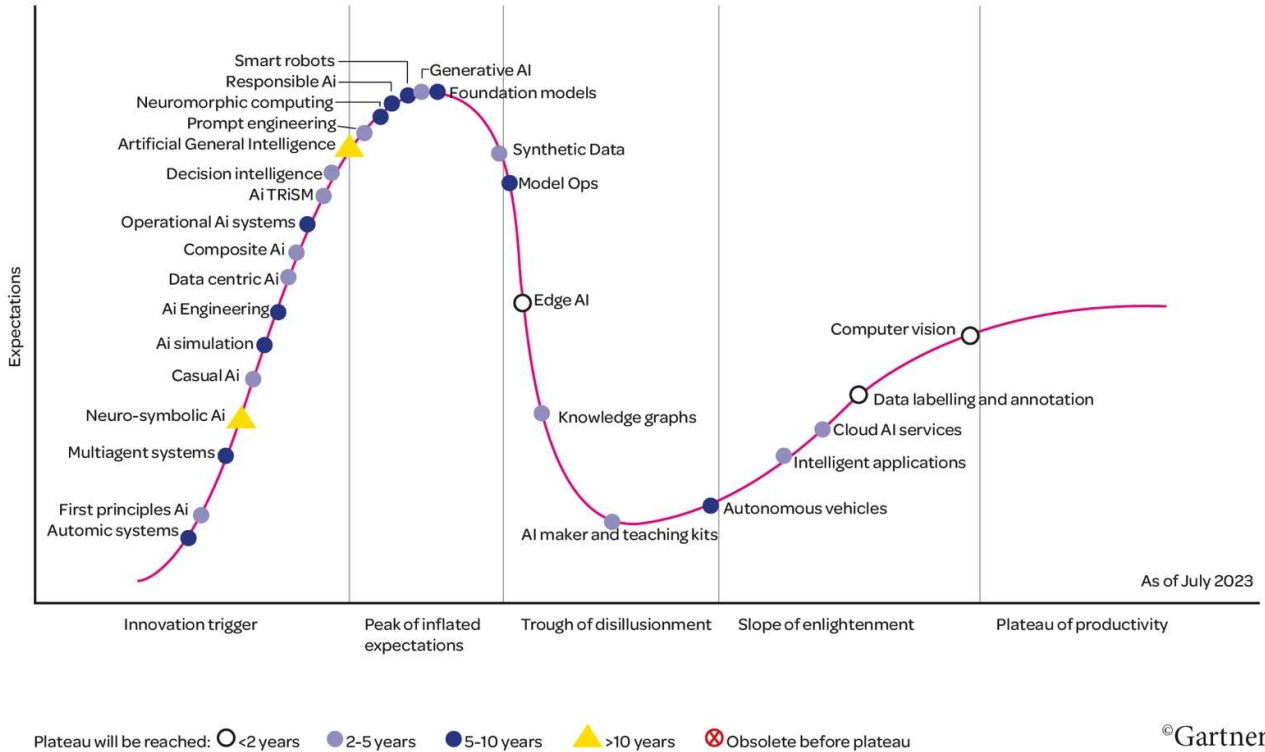
Synthetic data, therefore, has several advantages over real data:

- Privacy is enhanced because no actual PII is included in the set.
- Small segments and/or hard to reach groups can be extrapolated to boost sample sizes for analysis.
- Well-trained models show strong correlations with real-world data sets that they are based on

However, we believe claims that synthetic data will completely replace primary data collection (as Ritson suggests) are unlikely for a few reasons:

- These claims are common when new technology becomes available and synthetic data is currently over-hyped.
- Synthetic data will always be retrospective and can never be current or show emerging trends, making it unsuitable for longitudinal research.
- People are complex, nuanced creatures and replicating that nuance with synthetic data is difficult with current technology.
- Synthetic data will replicate biases in its training data and algorithm.

- There are no currently accepted ways to evaluate the quality and accuracy of synthetic data, leading to understandable client scepticism.
- Ethical and regulatory concerns



## Market Research in the age of AI

The process of conducting primary research can be manual and complex, with numerous tools and suppliers at each stage and limited tool consolidation. Some audiences are also hard to measure at scale reliably (i.e. C-suite and Decision Makers).

As a result, four core challenges emerge in the industry:

1. Lack of capacity, consistency and research expertise
2. The process is complex, with significant time and resource investment.
3. Insights are often under-utilised and hard to access
4. Small but important groups can be misunderstood or not examined at all

Generative AI and synthetic data may help to alleviate some of these challenges and provide new opportunities for the industry:

- Creating synthetic data to:
  - Explore edge cases that are uncommon in most datasets.
  - Enhance data sets by adding unmeasured variables.
  - Bias mitigation e.g. trying to mitigate bias in real world data to lead to fairer outcomes
  - Create a sandbox for scenario testing and simulations.
- Using AI research assistants to automate manual processes and add a layer of consistent interpretation of research data.
- Artificial survey/interview agents to conduct data collection with limited human involvement.
- Creating synthetic customers for persona's within an LLM to answer questions or review content
  - Here is a [basic version](#) of that functionality.

## Strategic framework - application and commercialisation

### In house Research Applications

Synthetic data can make conducting some types of research easier for brands. We see the most immediate and valuable potential uses are creating synthetic customers, sandbox environments, and edge case data sets. They will allow brands to leverage their current data to develop tools for internal review and exploration of blue sky areas. For example, opportunities could include:

1. **Scaling creative testing** in a quick, efficient and inexpensive way. Synthetic data could be a great way to test ads at scale.
2. Using synthetic data to **deepen understanding of small and niche audiences or**; such as a small set of business decision-makers for b2b brands or vulnerable audiences to avoid over researching.
3. AI research assistants could help create efficiencies and ultimately **speed up work**, allowing Insight professionals to more consistently and scalably use self-serve research applications (this would have knock on impact on agencies).

### Agency Research Applications

For agencies, we believe that AI's power lies in its ability to process vast datasets rapidly, uncover hidden patterns, and offer real-time insights that are beyond human capabilities alone. Whereas synthetic data at its

core, is designed to mimic the statistical properties of real data and enables researchers to operate across broader geographical regions, delve into new market segments or assists us in understanding under-represented groups more easily.

1. Improve accuracy and reliability of data, negate researcher biases
2. Find trends and patterns in large scale quantitative data and allow us to analyse qualitative insights at scale.
3. as well as find general efficiencies throughout the research process

### A note on potential limitations and getting the most out of AI

Whilst AI can speed up many boring tasks, to get the most out of it, it needs to be fine-tuned by human expertise. The research industry needs to remain alert as the use of AI and synthetic data does not come without limitations; the very essence of qualitative research (human empathy and understanding) remains irreplaceable, AI isn't quite at the point where it can fully understand project logic and context (though it is getting better all the time), some AI models can hallucinate and provide incorrect or fabricated responses and we must consider the upcoming emerging talent in the industry (could industry new-comers become too reliant on AI and treat what it says as gospel?).

By leveraging AI and synthetic data responsibly, research agencies can not only enhance operational efficiency but also drive innovation in exploring new methodologies for data collection and analysis.

### A strategic framework for learning

Ultimately, we see the below potential applications for AI in the Market Research Industry. over the course of the next 5 months, the MRS Advanced Insight & Analytics Council will be producing a guide to each of these applications to be used as a tool by in house and agency researchers alike.

	Current application	Areas for future exploration
Ai research		
Synthetic customer bots		
Ai survey/interview agents		
Synthetic data		
Turnkey research platforms		
Panels		
Research companies		
Custom LLM models		

## Appendix

Examples of companies operating in this space are:

### **AI research assistants**

- <https://www.taskade.com>
- <https://www.lucy.ai>
- <https://www.perplexity.ai>

### **Synthetic customers**

- <https://www.syntheticusers.com>

### **AI Survey/interview agents**

- <https://www.delvinia.com/solutions/cris/>
- <https://www.nexxt.in/inca/platform>
- <https://groupsolver.com>

### **Synthetic data**

- <https://hazy.com>
- <https://mostly.ai>
- <https://www.evidenza.ai>
- <https://gonative.ai/post/synthetic-data-for-market-research>

### **Turnkey platforms**

- <https://www.quantilope.com>
- <https://www.kantar.com/marketplace>
- <https://www.zappi.io/web/>
- <https://conjointly.com>
- <https://www.goideally.com>
- <https://www.qotracksuit.com/au>
- <https://sightx.io/about>

### **Panel companies**

- <https://www.newtonx.com>
- <https://www.mindforceresearch.com/online-sampling>
- <https://www.purespectrum.com>
- <https://www.lightspeedresearch.com.au>
- <https://tolunacorporate.com>
- <https://www.dynata.com>