



Advanced Data Analytics group

AI and Synthetic Data in Market Research

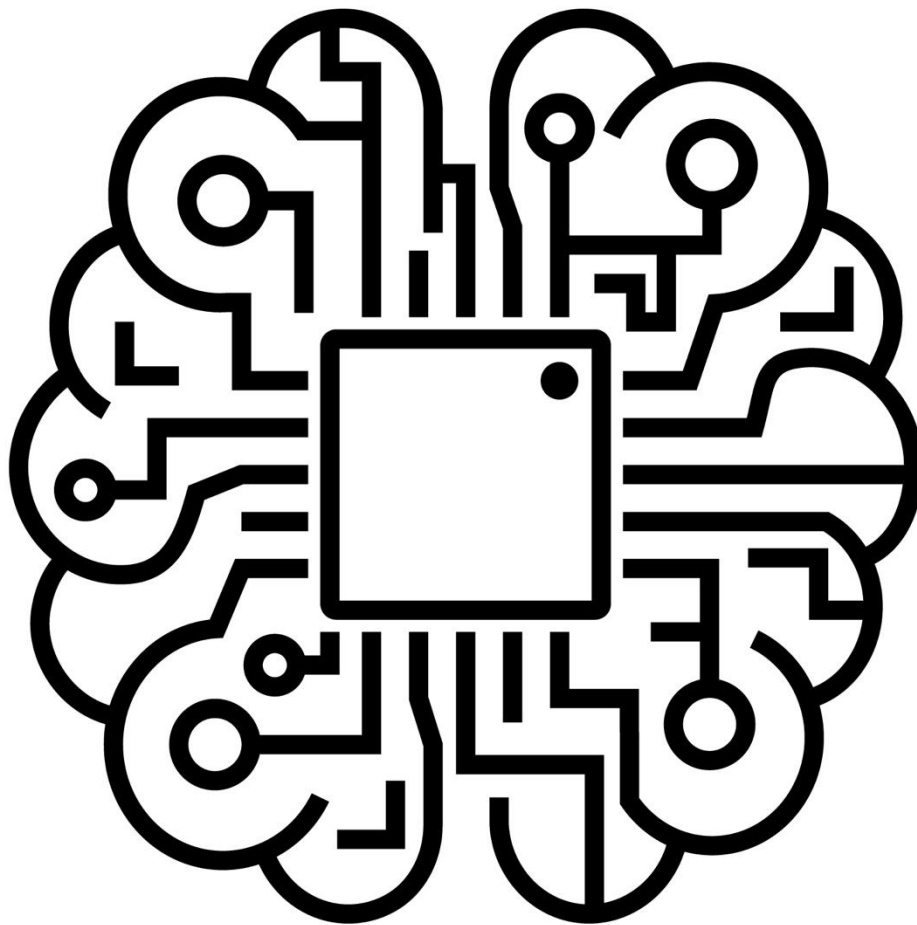
Authors:

Sam Barton

Jarrold Howard-Payne

Anthony Shephard-Williams

Jake Steadman



About the MRS Advanced Insights & Analytics (AIA) Council

The **MRS Advanced Insights & Analytics (AIA) Council** is a specialist group within the **Market Research Society (MRS)**, dedicated to advancing the use of cutting-edge data science, AI, and analytics in market research. Our mission is to help research professionals navigate the fast-evolving landscape of AI-driven insights by fostering industry collaboration, setting best practice standards, and driving innovation.

The Council comprises industry leaders from global brands, research agencies, and academic institutions, working together to define the future of market research. Through white papers, thought leadership, and practical frameworks, we aim to shape a more informed, responsible, and effective use of AI and synthetic data in market research.

Introduction: The Synthetic Data Hype vs. Reality

The promise of synthetic data is exciting, but its ability to revolutionise research remains a contested question. In our view, the industry needs to move past the hype and focus on its practical, high-value applications.

Synthetic data presents undeniable advantages -particularly in privacy compliance and scalability - but its real value is more nuanced. Instead of replacing real-world data collection, we believe synthetic data should be viewed as an enhancement tool that fills gaps, stress-tests assumptions, and refines research methodologies rather than as a standalone replacement.

The key question isn't what synthetic data *can* do—it's what it *should* do. This paper explores its practical applications, challenges over-hyped claims, and lays out a more grounded vision for its role in market research.

What Synthetic Data Actually Offers (and What It Doesn't)

Synthetic data is artificially generated, modeled to resemble real-world datasets. While many claim it can replace primary research, we believe its true power lies elsewhere: in supplementing insights, expanding testing capabilities, and refining predictive analytics.

Key Advantages (That Actually Matter)

- **Plugging Data Gaps, Not Replacing Data:** AI can model edge cases or generate data where real-world collection is infeasible, but it should never replace real human interactions.
- **Privacy Without the Trade-Offs:** Synthetic data allows researchers to work without exposing sensitive PII. However,

privacy should never come at the cost of authentic behavioral insights.

- **Scenario Simulation Over Static Analysis:** The best use of synthetic data isn't to recreate reality but to simulate possibilities, stress-test business decisions, and uncover counterintuitive trends.
- **Expanding Sample Sizes for Hard-to-Reach Audiences:** For niche or sensitive groups, synthetic data can provide additional robustness to insights, but it must be validated against real-world findings. Similar to data weighting—though not identical in function—synthetic data can help address small sample sizes by generating additional data points rather than merely adjusting weights to compensate for imbalances.

Misconceptions About Synthetic Data

- **Synthetic Data Is NOT Bias-Free:** AI is trained on biased datasets, and synthetic data inherently inherits those biases. If not carefully managed, synthetic data can amplify rather than mitigate bias in market research, leading to misleading conclusions.
- **Bias in Model Development:** An often-overlooked source of bias is the programmers who build AI models. Their perspectives, assumptions, and unconscious biases can influence the code itself, embedding unintended biases into the model's logic and decision-making processes.
- **Inability to Capture Emerging Trends:** Synthetic data is derived from historical patterns and lacks real-time adaptability. As a result, it struggles to reflect sudden consumer behavior shifts driven by cultural changes, economic crises, or viral trends. Recent examples include pre- and post-COVID travel behavior and shifting attitudes toward mental health.
- **More Data ≠ Better Insights:** There is a false assumption that generating more synthetic data leads to better decision-making. In reality, excessive reliance on synthetic data often results in overfitting and misleading patterns, reinforcing outdated or incorrect conclusions rather than uncovering new insights.

Our Take: The Right Way to Use Synthetic Data

The market research industry is at a crossroads. Rather than embracing synthetic data as a universal fix, we argue for a more strategic approach, using it as a precision tool rather than a blunt instrument.

Application	How It Works	Our Opinion	Risks
Enhancing Behavioral Models	AI-generated personas simulate different consumer responses.	Useful for preliminary testing but should always be validated against real-world data.	May oversimplify complex behaviors, leading to misleading insights.
Filling Gaps in Longitudinal Research	Simulating consumer sentiment shifts over time.	Promising, but dangerous if assumed to be real-time consumer behavior.	Backward-looking, making it unreliable for detecting emerging trends.
Creative & Ad Testing at Scale	Simulating ad reactions before real-world deployment.	Strong use case, but only if real consumer feedback is used alongside it.	Lacks emotional nuance and struggles to detect cultural misalignment, risking ineffective campaigns.
Bias Mitigation	Training AI to recognise and counteract biased datasets.	Conceptually sound but dependent on human oversight—AI alone can't 'fix' bias.	Can still reinforce hidden biases if not properly managed.
Product Development & Innovation	Synthetic data simulates customer needs, preferences, and usage patterns to guide product design	Useful for brainstorming and prototyping, but real user testing is essential before launch.	May not fully capture real-world variability, leading to products that miss key consumer pain points.
Customer Experience (CX) Simulation	AI-generated personas simulate customer interactions across touchpoints to improve UX design	Helpful for stress-testing digital experiences, but real customer feedback should validate insights.	May oversimplify emotions and complex decision-making processes, leading to misleading recommendations.

In-House vs. Agency Use Cases

Synthetic data has distinct applications depending on whether it's used in-house by brands or within research agencies.

In-House Research Teams

- **AI-Powered Self-Serve Insights Platforms:** Brands can integrate synthetic data into their own insight tools to enhance agility and decision-making without relying on external research suppliers.
- **Scaling Creative and UX Testing:** In-house teams can use synthetic consumers to rapidly test advertising, messaging, and product UX before large-scale launches.
- **Market Expansion Analysis:** Synthetic data can help brands predict potential consumer behavior in new geographic markets before investing in real-world market entry studies.

Research Agencies and Consultancies

- **Automating Large-Scale Data Processing:** AI-generated datasets can enhance trend forecasting, competitive intelligence, and segmentation work for clients. However, agencies must carefully navigate the ethical implications of using client-owned data to train AI models. The question of data ownership and reuse remains a grey area, as client data is typically proprietary and not meant for broader application.
- **Dynamic Scenario Modeling for Clients:** Agencies can offer AI-driven simulations to help brands anticipate customer reactions in different business conditions. Synthetic data could play a role here by allowing agencies to create proxy datasets that mimic real-world data without breaching ownership constraints. While this approach offers a potential solution, it does not fully resolve the ethical concerns surrounding data usage rights.
- **Ethics & Bias Consultancy:** Agencies have an opportunity to position themselves as leaders in ethical AI application by auditing synthetic datasets, ensuring fairness, and guiding clients on responsible data use. As AI adoption grows, agencies must establish clear guidelines on data provenance, synthetic data usage, and transparency to build trust and accountability in AI-driven insights.

What Needs to Change in the Industry?

The enthusiasm around AI-driven research is understandable, but in our opinion, many organisations are misapplying synthetic data in ways that could backfire. We propose the following shifts in thinking:

1. **Stop Framing Synthetic Data as a Replacement for Real Data:** It's a complementary tool, not a standalone solution.
2. **Prioritise Explainability Over Black-Box Modeling:** Researchers should understand and validate how synthetic data is generated before relying on it for decisions.
3. **Improve Transparency in Synthetic Data Provenance:** Organisations should disclose how synthetic data is created, what real-world data it is based on, and any assumptions made in its generation. Clear documentation and transparency will help build trust and ensure responsible usage.
4. **Develop Clear Standards for Synthetic Data Quality:** Currently, no widely accepted benchmarks exist for evaluating synthetic data accuracy—this must change.
5. **Use AI to Stress-Test, Not Replace, Human Judgment:** AI's strength lies in generating hypotheses, not making final decisions.

Looking Ahead: The MRS AIA's 2025 AI Initiative

Throughout 2025, the MRS Advanced Insight & Analytics (AIA) Council will publish a series of deep-dive papers exploring both the opportunities and limitations of AI-driven research methodologies.

We'll be tackling:

- **The Future of AI-Augmented Research Teams:** How will the role of researchers evolve in an AI-driven world?
- **AI and the Ethics of Synthetic Data:** How do we ensure transparency, fairness, and accountability?
- **Beyond Synthetic Data: Next-Generation AI Applications:** What's next for AI-driven insights, and how can we prepare?

We encourage market researchers, data scientists, and insights professionals to engage with this series, challenge our perspectives, and contribute to the debate. The future of AI-driven market research isn't set in stone, it's something we must shape together.

Conclusion: Our Vision for AI & Market Research

AI and synthetic data will undoubtedly play a major role in market research's future, but perhaps not in the way currently assumed by some. The real opportunity lies in using AI-driven data responsibly, as a tool for scenario-testing, hypothesis generation, and augmentation, not as a replacement for direct consumer research.

The future won't be about choosing between synthetic and real data, it will be about mastering their interplay. Those who strike this balance will lead the industry into the next era of insights-driven decision-making.