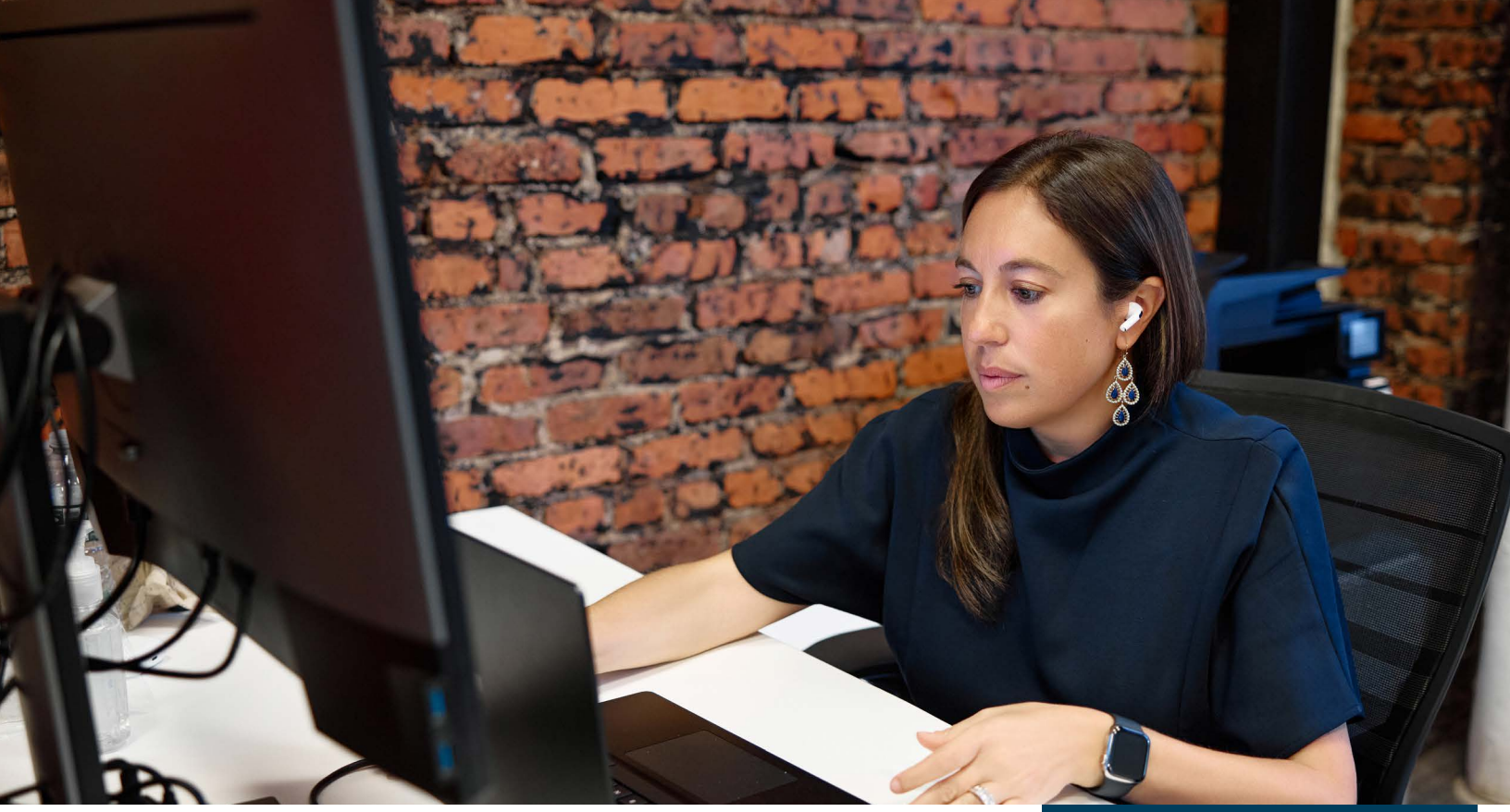

Filling the Gaps

How Synthetic Data is Reshaping the Marketing Landscape

arima

Executive Summary

The marketing and advertising industry has reached a structural inflection point. While the volume of available data has increased exponentially, marketers face diminishing visibility into the full customer journey due to tightening privacy regulations, ecosystem fragmentation and the rise of closed platforms. As a result, even the most sophisticated data infrastructures struggle to produce a complete, actionable view of the market.

Synthetic data has emerged as a pragmatic, privacy-first solution to these challenges. By generating statistically representative population datasets derived from multiple real-world inputs, synthetic data enables organizations to model entire markets—not just observable users—while avoiding reliance on personally identifiable information (PII).

This paper outlines the role synthetic data can play across the modern marketing workflow, explains its underlying methodology and highlights its practical applications in audience strategy, measurement and planning.

> “The main benefit of utilizing synthetic data is to fill gaps — gaps in the data we use, the customer insights we’re looking for, as well as in our understanding of the marketplace. The question is not whether your deterministic data is accurate; it is what you are systematically missing because of its inherent limitations.”
>
> Winston Li, Founder, Arima

The Structural Problem: Incomplete Visibility in a Data-Rich World

Despite significant investment in data partnerships, identity resolution and clean room environments, most organizations operate with structurally incomplete datasets. Three persistent challenges define this reality:

01 Fragmentation

Data ecosystems are inherently incomplete. Even well-integrated systems fail to fully represent the total addressable market, systematically under-representing prospects and non-customers.

02 Inconsistent Data Freshness

Data sources vary widely in update cadence and methodology. This creates temporal misalignment, where datasets quickly become outdated or inconsistent when combined.

03 Cost and Operational Complexity

Building and maintaining comprehensive data infrastructure requires substantial and ongoing investment in engineering, licensing and governance.

The net effect is a paradox: organizations often have deep insight into existing customers but limited understanding of the broader market where future growth resides.

What Is Synthetic Data?

Synthetic data, in a marketing context, is a statistically representative model of a real-world population constructed from multiple data sources. It does not replicate actual individuals. Instead, it uses generative modeling techniques to preserve the statistical properties of source data while simulating realistic individuals and households.

These datasets are typically built using inputs including:

- Census and demographic data
- Market research surveys
- Mobility and location intelligence
- Points-of-interest data
- Transaction and purchase behavior

The resulting synthetic population enables marketers to analyze behavior, simulate outcomes and explore market dynamics at scale—without directly processing PII.

Key Properties

Total Market Coverage	Synthetic datasets model entire populations within defined geographies, enabling unbiased analysis of users and non-users alike.
Dynamic and Refreshable	Unlike static panels or periodic surveys, synthetic populations can be continuously updated as underlying data sources evolve.
Privacy by Design	Because synthetic individuals are simulated rather than matched, datasets are inherently privacy-safe and well-suited for regulated industries.
Customizable	Organizations can layer proprietary data—such as CRM segments, brand studies, or behavioral signals—onto synthetic populations to tailor them to specific business needs.

Synthetic Data vs. Traditional Approaches

Traditional marketing analytics approaches each suffer from structural limitations including but not limited to:

- Panel data provides depth but lacks scale
- Lookalike modeling is constrained within platform ecosystems
- Marketing Mix Modeling (MMM) operates at an aggregate level, limiting granularity

Synthetic data bridges these gaps by enabling full-population modeling with individual-level simulation across channels. It combines the breadth of MMM with the granularity of user-level analysis—without relying on deterministic identity matching.

> “One of the most promising aspects of synthetic data is its ability to connect planning and measurement more closely—bringing together scale, granularity and simulation in a way that traditional approaches haven’t been able to.”
>
> Kevin Freemore, 4As

Implementation Framework

Adopting synthetic data does not require replacing existing systems. The most effective implementations follow a phased approach:

01 **Phase 1: Initial Testing**

Organizations begin with out-of-the-box synthetic datasets to run baseline simulations and exploratory analyses. These outputs serve as directional insights and are often used alongside existing methodologies.

02 **Phase 2: Customization**

Proprietary data sources—such as first-party behavioral data or brand-specific research—are layered into the synthetic population to improve calibration and relevance.

03 **Phase 3: Integration and Scale**

Over time, synthetic datasets can be embedded into planning and measurement workflows, supporting ongoing simulation, forecasting and optimization efforts.

This progressive model reduces risk while enabling rapid demonstration of value.

Applications Across the Marketing Workflow

Audience Strategy and Segmentation

Synthetic data enables marketers to define and analyze audience segments across the full market, including individuals not present in CRM or first-party datasets. This expands strategic visibility and improves prospecting accuracy.

Measurement and Marketing Mix Modeling

Synthetic populations allow for the construction of simulated consumer journeys, linking exposures, behaviors and outcomes at the individual level. This enhances traditional MMM by introducing greater granularity while maintaining cross-channel scope.

Simulation and Scenario Planning

One of the most powerful applications of synthetic data is the ability to run "what-if" scenarios. Organizations can model the impact of changes in targeting, creative, channel mix, or geography before making any investments.

Validation and Benchmarking

Synthetic populations provide an independent reference model for evaluating media performance. This enables marketers to benchmark platform-reported outcomes against modeled expectations in a privacy-safe manner.

Case Illustration

A large agency network working with a quick-service restaurant (QSR) brand integrated synthetic population data with proprietary research and real-time polling inputs. The resulting customized dataset was used across planning, analytics and activation workflows.

By incorporating mobility data into the synthetic population, the team was able to enhance geographic targeting and improve coordination between strategy and execution. Outputs were also integrated into the broader technology stack for reporting and downstream analysis.

> "The value of synthetic data isn't in replacing what agencies already have, it's in strengthening it. When used thoughtfully, it becomes a layer that brings context, scale and perspective to existing data investments."
>
> Amy Pacheco, 4As

Strategic Implications

Synthetic data should not be viewed as a replacement for first-party data or deterministic identity systems. Instead, it functions as a complementary layer that addresses structural blind spots inherent in all observable datasets.

Organizations evaluating synthetic data may consider the following approach:

Reframe the Question

Focus less on whether synthetic data matches existing datasets and more on what existing datasets fail to capture.

Start with a Defined Use Case

Prioritize targeted applications such as audience sizing, geographic planning, or pre-campaign simulation.

Invest in Iterative Calibration

Treat synthetic populations as evolving assets that improve over time through continuous refinement and validation.

Conclusion

As the marketing ecosystem continues to shift toward privacy-first architectures and fragmented data environments, the ability to model the unseen portions of the market will become increasingly critical. Synthetic data can create a scalable, privacy-compliant pathway to achieving this visibility.

By enabling full-market representation, enhancing measurement frameworks and supporting forward-looking simulation, synthetic data positions would be another tool for organizations to make more informed, resilient decisions in an increasingly uncertain landscape.

About the 4As

The 4As was established in 1917 to promote, advance and defend the interests of our member agencies, employees and the advertising and marketing industries overall. We empower and equip our members to confidently navigate the ever-changing ecosystem of the agency world. We ensure they remain relevant, are positioned to compete and have the resources to thrive and grow. With a focus on advocacy, talent and creating impact, the organization serves 600+ member agencies across 1,200 offices, which help direct more than 85% of total U.S. advertising spend. The 4As includes the 4As Benefits division, which insures more than 160,000 employees; the government relations team, who advocate for policies to support the industry; and the 4As Foundation, which advocates for and connects rising talent to the marketing industry by fostering a culture of curiosity, creativity and craft to fuel a more equitable future for the industry.

About Arima

Founded in 2020, Arima is a data and software development company, supporting agency and marketer workflows through tools such as Persona Insights, Cross Media Planner, unlimited-use MMM and InstaPoll.

Arima solutions are powered by Synthetic Society™, a statistical recreation of the real population (from national to hyper local), capable of simulating consumer perceptions, behaviors and test marketing and media effectiveness before a single dollar is spent.