

Kana Synthetic Data

For Brands and Marketers: Audience Expansion and Augmentation

The Challenge

Brand and marketing teams hit the same wall twice: the highest-quality first-party data is too small to model against, and the largest data they can buy or receive is too thin to activate.

- First-party signals (loyalty, DTC, panel, and sign-up data) are deep but small. Too few records to train robust models, run reliable lookalikes, or stage realistic activations.
- Retailer feeds, syndicated panels, clickstream, and media-partner files are large but anonymous, with no demographics, lifestyle attributes, or value tier attached to the behavior.
- Third-party datasets are expensive and inconsistent at \$0.50–\$2.00 per record across millions of records yet the same user can appear as both male and female in the same file.
- Lookalike models built on thin first-party seeds or noisy third-party data produce unreliable segments, wasting media spend on experimentation that doesn't convert.

The result is a compounding cost problem: wasted media spend on poorly targeted campaigns, missed launch windows in high-growth markets where no usable third-party data exists, and continued dependency on data vendors whose quality degrades the very ROI marketers are trying to defend.

Solution Overview: Kana Synthetic Data Generation

The Solution

Kana's Synthetic Data Agents solve both problems with one unified solution. Audience Expansion Agents turn a small, trusted seed into a faithful population at the size downstream workflows actually need. Audience Augmentation Agents fill in missing attributes on a large entity file by predicting them directly from the behavior already in the record.

Both produce datasets, entities and events shaped like first-party data, without reproducing any individual seed record and without third-party identity graphs.

How It Works

Audience Expansion: scale a trusted seed into a faithful population

- Ingests a small entity-and-events seed from existing Kana datasets, an uploaded CSV pair, or an AI-generated description.
- Trains a new per-dataset Kana generative model on the seed, encoding mixed-type customer attributes into a smooth, sampleable latent space that preserves the joint distribution.
- Samples a synthetic population whose marginals and correlations match the seed, then synthesizes per-customer interaction events to match.
- Outputs a real entity table plus a real interaction-events table at 50K, 500K, or 5M scale, drop-in for segmentation, modeling, lookalike training, and reach planning. No individual seed record is reproduced.

Audience Augmentation: predict missing attributes from behavior

- Takes a large entity file plus its interaction events (purchases, views, visits, impressions) and the list of attributes you need: life-stage, value tier, channel preference, household income, category affinity, primary need-state, and more.
- By default, builds calibrated per-item attribute distributions via domain-aware LLM research, then mixes them per entity weighted by behavior.

www.kana.ai



Solution Overview: Kana Synthetic Data Generation

- Switches automatically to a supervised gradient-boosted-tree ensemble when a labeled training seed is available, same probability output, drop-in compatible.
- Outputs the original entity table augmented per attribute with an argmax prediction, a calibrated per-value probability vector, and a confidence score so your activation team can threshold on what to trust.
- Run back-to-back with Expansion as a self-evaluation: the engine predicts attributes it never saw, so you can compare to ground truth and prove accuracy on your own data before any roll-out.

How You Benefit

- Stand up realistic populations in any market, including ones with zero usable third-party data, without exposing real PII.
- Activate large anonymous files that previously couldn't be used, by predicting demographics and lifestyle attributes directly from the behavior already in the record.
- Reduce or eliminate third-party data spend — at \$0.50–\$2.00 per record across millions of records, that is real money — while improving targeting precision and audience overlap.
- Pre-flight experimentation: simulate thousands of segment-by-creative-by-budget combinations against the synthetic population to predict winners before any media is spent, replacing live A/B burn with offline simulation.
- Hand finished entity and events tables directly to media planning, modeling, measurement, and brand teams — no further data engineering required.
- Push attribute predictions and probabilities directly into Meta, Google, TikTok, YouTube, TradeDesk, and your DSP/CDP of choice using your own hashed-email identity key.
- Move from raw seed to activation-ready dataset in minutes, not weeks.



Solution Overview: Kana Synthetic Data Generation

Why Kana?

Here are a few more ways in which we're proud to be unique:

- **Agentic AI Built by MarTech Architects:** Our founding team has spent decades at the intersection of AdTech and MarTech.
- **Data Governance and AI Policy Management:** Centralized data governance and policy enforcement provides a layer of trust and satisfies AI security audits.
- **Insane Velocity Provides You With Speed to Action:** We deploy instantly into your existing stack and move you from raw data to revenue-generating insights in minutes.
- **No Assembly Required:** We eliminate the technical tax of AI by providing the builders, the logic, and the execution layers as a unified service.
- **Future-Proof Architecture:** As your business changes, the agentic layer evolves to meet new challenges.

www.kana.ai

