



Powering AI Data Agents with Synthetic Behavior Data

Conviva × Rockfish · Customer Story

At the ACM [CAIS 2026 conference](#), Conviva unveiled **NEXA**, an **AI Analytics Agent** that brings *user-oriented* analytics to the agent era. NEXA starts from the user and answers the questions the business actually cares about: **did the agent achieve the user's intended outcome and business impact, where was it not effective, and why**. It delivers those answers two ways — automatically, surfacing business-impacting insights that would take human analysts hours or days (or get missed entirely), and as an interactive Agent an analyst can interrogate directly to drive an investigation. Because NEXA reasons across both app behavior and chat conversations, demoing its depth required realistic e-commerce and shopping-chat data: Conviva's product director Haijie Wu needed data "inspired by true stories" but free of real customer data — rich enough to carry the specific behavioral signals and labeled failure modes that show NEXA at its best.

About Conviva. Conviva is the leading real-time streaming and experience analytics platform. Its patented Time-State Analytics processes trillions of events a day for businesses such as Disney, Salesforce and MLB — providing product- and user-level visibility and insights. That same real-time, pattern-detection capability is now extending to AI agents, the context for this work. www.conviva.ai

About Conviva NEXA

NEXA is built to catch the user's intention, sentiment and outcome that matter most. One real customer example: users who visited the pricing page before asking the chat agent product questions showed higher intention to purchase, and converted more often. That insight lives at the intersection of digital behavior and agent interaction; clickstream, chat traces, or backend metrics alone would each miss it. The demo data had to reproduce exactly this kind of cross-surface pattern — or it couldn't show NEXA's value.

A high-intent journey makes the point — the pricing-page visit and the chat that follows are one connected story (*generated example, abridged*):

WEB EVENTS viewed product detail page → viewed pricing page → opened chat

user (high intent): Does the Pro plan include priority support and extra seats?

agent: Yes — Pro covers priority support and up to five team seats.

user (high intent): Perfect, I'll upgrade now.

→ **outcome: purchased**

The Challenge

Detection logic is only as trustworthy as the data you test it on. The insights NEXA surfaces live at the intersection of two streams — what a user does in the app and what they say to the chat agent. But real production data can't be shared for privacy reasons and tends to lack clean examples of the rare patterns that matter most. Conviva's team tried prompting an LLM and hand-building simulators; both proved cumbersome and low quality. They needed a dataset that connected web behavior to chat utterances, with the target patterns present and explicitly labeled, as can be seen in Conviva's real production customer datasets.

The Rockfish Approach

Rockfish helped Conviva build a custom, expressive data-generation pipeline — on its **AgentFuel** tooling — to demo and test NEXA on a short turnaround. The core idea: define custom **personas**, each a distinct behavioral profile with its own finite-state-machine transitions for web browsing and chat, then generate a mix of personas with intended outcomes to stress-test the agent. Everything downstream flows from the persona, which keeps the patterns coherent rather than bolted on.

1. **Structured generation** — one workflow per persona produces an entity-relational schema: users (who they are), sessions (each visit), web events (the clicks within a visit).
2. **Cross-stream bridge** — keeps web and chat consistent: a chat only exists if the user actually started one on the site, and only after they'd clicked for support.
3. **Pattern injection** — overlays the named failure modes: derives per-row columns, applies the A→B→A anchoring overlay, and compresses frustration-spiral timestamps to ~30s spacing.
4. **Persona-aware transcripts** — an LLM fills in realistic chat conversations turn by turn, matched to each persona's tone and state.
5. **Labeling & post-processing** — derives ground-truth detection flags and merges everything into a unified, time-sorted event log.

The pipeline was proven on one signal first (v1 — the intent pattern above), then generalized to the full set of labeled failure modes from Conviva's brief (v2), adding a product catalog and a dedicated pattern-injection stage.

Results

The planted intent signal reproduced cleanly. Across 1,000 synthetic users, pricing-page visitors behaved exactly as the real-world pattern predicted — at roughly 3× the rate of non-visitors:

Persona behavior	Pricing visitors	Non-visitors	More likely
Asked a product question	45.2%	14.5%	~3.1×
Requested a human agent	38.5%	10.6%	~3.6×
Purchased	40.1%	11.6%	~3.5×

n = 1,000 synthetic users. The intended signal is reproduced verifiably, not merely plausibly.

The four named failure modes were injected and labeled at scale — 1,000 users, 250 per persona, every instance flagged so detection accuracy can be scored against truth:

Pattern (failure mode)	Description	Prevalence in dataset
Frustration spiral (refund)	Refund auth-error spiral; agent only retries; user quits	52.5% 196 / 373 sessions
Sale-link → repeat stockouts	Arrives via sale link, hits 2+ stockouts within 5 min	5.0% 50 users
Anchoring → add-to-cart	A→B→A comparison, returns to A and carts it	11.9% 119 users
Multi-agent contradiction → abandon	External vs. storefront agent conflict, session abandoned	7.7% 77 users

Patterns Span Event and Conversation

The frustration-spiral pattern starts in the web events and escalates through the chat — which is exactly the cross-surface signal NEXA is built to catch (*generated example, abridged*):

e cross-surface signal NEXA is built to catch (generated example, abridged):

WEB EVENTS viewed order history → clicked “request refund” → opened support chat
user (calm): Hi, I’d like to request a refund for order #847291 — I was charged twice.
agent (calm): I’m sorry, I’m unable to verify your account — I’m getting an authentication error...
user (frustrated): This is ridiculous. I’ve given you everything three times now...
→ outcome: quit

The Impact

~3.5x behavioral lift reproduced cleanly in synthetic data	4 named failure patterns injected, labeled, and measurable	< 1 Week from a single proven signal to the full pattern set
--	--	---

- **Days, not weeks** — persona design to a production-quality, demo-ready dataset in days.
- **Structure + language together** — Rockfish for relational and temporal structure, an LLM for believable transcripts — neither alone produces a dataset that is both join-correct and conversational.
- **Labeled and measurable** — every pattern flagged per row, turning the dataset into a real detection benchmark, not just realistic-looking data.
- **Repeatable & private** — fully synthetic, no real customer data, and regenerable as NEXA evolves.

What's Next

- **Sharper agent testing** — run NEXA against the data blind to the labels and score discovery against ground truth.
- **Scale & extend** — regenerate at larger volumes and add new failure modes on demand.
- **Standardize the loop** — a repeatable benchmark Conviva reruns as NEXA evolves.

Why Rockfish Data

For teams building AI applications, Rockfish delivers enterprise-grade custom data that goes well beyond simple randomization:

- **Schema-aware generation** — preserves relationships, distributions and domain logic, not just plausible-looking values.
- **LLM integration** — semantically rich unstructured content (like chat) alongside structured tabular data, in one pipeline.
- **Pipelines that version with your product** — regenerate the data as your schema, features or use cases evolve.
- **Privacy-safe by design** — no real user data required, no compliance risk.

About Rockfish Data. Rockfish Data, built on foundational research from Carnegie Mellon University, generates high-coverage, labeled datasets purpose-built for evaluating AI models and data analytics agents in production. Unlike generic benchmarks, Rockfish systematically covers edge cases, rare scenarios and domain-specific failure

modes — giving teams the evaluation infrastructure to measure accuracy and reliability before deployment.
www.rockfish.ai