CASE STUDY

Nielsen Operationalizes Trust for a Production Multi-Agent AI Copilot

Nielsen

INDUSTRY

Media and Consumer Research

DEPLOYMENT TYPE

SaaS

AI OBSERVABILITY SOLUTIONS

- Agentic Observability
- Fiddler Guardrails

TECH STACK

- Google Cloud Platform
- Google BigQuery
- MLFlow
- Grafana
- Prometheus
- Fiddler AI

Nielsen is a global leader in audience measurement, data, and analytics, powering decision-making for media companies, advertisers, and brands worldwide. As part of its AI transformation strategy, Nielsen launched 'Ask Nielsen'; a multi-agent AI analytics copilot that enables users to query complex datasets conversationally and receive business-ready insights in real time.

Results: Enterprise-Grade Safety Without Sacrificing User Experience

- **Jailbreak Defense:** 95.38% accuracy and 99% precision detecting direct harmful instructions; 98% accuracy blocking persona-play attacks. Outperforms Bedrock Guardrails.
- **Brand Safety:** 100% accuracy and precision filtering toxic content across violence, hate speech, profanity, sexual content, and misconduct.
- **Low-Latency Enforcement:** Synchronous guardrails at <100ms latency, with no impact on user experience.
- **Root Cause Resolution:** Step-level diagnostics pinpoint degradation within specific agents, reducing time to resolution.
- **Safe Scaling:** Consistent runtime enforcement maintained as the system expanded across new datasets, client segments, and agent types.

The Challenge: Scaling Multi-Agent AI Without Compromising Trust

'Ask Nielsen' is not a simple chatbot. It is a multi-agent orchestration system that includes a master routing agent, domain-specific agents for Campaign Analysis, Competitive Intelligence, Audience, and Content, and sub-agents for natural language processing, planning, SQL generation, code execution, and insight narration.

'A single user request may trigger multi-step reasoning, SQL query generation, Python code execution, data visualization, narrative insight generation, and real-time guardrail validation. In a representative trace, that means over 1,000 spans, 100+ LLM calls, and 50+ guardrail calls.

As 'Ask Nielsen' scaled, the team identified two categories of practical risk that had to be addressed before the system could serve enterprise clients reliably.

1. **Governance and Security Risk:** Customer-facing agents increase the operational blast radius. Even when underlying data is trusted, agent behavior can drift outside intended boundaries, particularly under adversarial pressure. Direct and indirect jailbreak attempts, where adversaries try to bypass safety filters and push the system into unsafe behavior, posed a concrete threat.
2. **Brand Safety Risk:** Toxic content surfacing in generated responses creates reputational harm if not filtered reliably. In parallel, the team needed to preserve a high-quality end-user experience, meaning guardrail enforcement could not introduce latency that made interactions feel slow or unreliable.

The objective was clear: ship fast without compromising safety, consistency, or compliance.

Traditional observability tools could surface traces, but not answer the harder question: which step in the multi-agent flow caused degradation, and why. Without that visibility, both risks remained unresolved until the point of damage. Nielsen needed an alternative: a runtime trust layer that could operate at the agent level, not just report on what had already gone wrong.

The Approach: A Runtime Trust Layer for Agentic AI

Rather than treating governance as a periodic review, Nielsen built for continuous oversight. Their framework emphasized three outcomes: keeping the system within intended operational boundaries in the face of adversarial behavior, maintaining brand safety across key toxicity categories, and improving quality and reliability through continuous evaluation and production observability, supported by actionable diagnostics.

Nielsen partnered with Fiddler AI to serve as the runtime trust and governance layer for 'Ask Nielsen', integrating directly into the multi-agent architecture alongside existing orchestration and telemetry systems.

Establishing a Unified Evaluation and Observability Framework

The firm uses Fiddler as a core component of their agentic solution by integrating Fiddler Trust Models for in-environment evaluation, and Guardrails to support safe and compliant LLM deployments. The team tracks 40+ metrics across reliability, performance, and business impact, each paired with diagnostics to support root cause analysis.

Nielsen operates a rigorous evaluation lifecycle built around golden datasets per primary agent, offline benchmarking, online monitoring using user feedback and baseline comparisons, and drift detection using clustering techniques. Fiddler bridges offline and online evaluation by continuously monitoring runtime performance and surfacing regressions when production metrics fall below validated baselines.

To make observability actionable, the team uses Fiddler's query clustering to identify patterns in user queries, diagnose behavior across similar query types, and proactively address clusters where the model may be underperforming.

Runtime Safety Enforcement with Real-Time Guardrails

'Ask Nielsen' enforces Fiddler Guardrails for Safety, PII, and Faithfulness at critical checkpoints within the LLM gateway layer, ensuring policy compliance before responses reach users. Guardrails operate with synchronous enforcement paired with asynchronous monitoring, delivering governance without degrading the speed of customer interactions.

Visibility Across the Agentic Hierarchy

'Ask Nielsen' exports OpenTelemetry traces for agentic workflows. Fiddler transforms these traces into actionable intelligence, enabling the team to isolate degradation to specific agent steps, understand whether issues originate in natural language processing, SQL generation, code execution, or insight narration, monitor guardrail enforcement across complex flows, and alert on completion rate and faithfulness drops. This step-level visibility allows teams to resolve issues faster and deploy improvements with confidence.

> "Fiddler delivered unified observability, protection, and governance across agents and predictive models, making it fundamental to our AI strategy. We are especially looking forward to the upcoming rollout of the control plane and the intelligent orchestration it will bring as our AI footprint grows."
>
> **Karthik Rao**, CEO, Nielsen

Looking Ahead

Nielsen continues to expand 'Ask Nielsen''s capabilities across additional datasets, client segments, and agent types. By pairing continuous evaluation with production observability, and enforcing safety in the runtime path, Nielsen built a system that is both usable and trustworthy. The result is a more interactive way for customers to access insights, backed by measurable governance and brand-safety performance.

With Fiddler as the runtime trust layer, Nielsen can innovate confidently, ensuring that as AI systems grow more powerful, they remain safe, governed, and reliable.

To learn more about the Fiddler Unified AI Observability and Security Platform, book a demo at fiddler.ai/demo or read additional at fiddler.ai/customers.

fiddler.ai sales@fiddler.ai