



Make Your Data Analysis Agents Reliable in Production

Rockfish AgentFuel helps enterprises evaluate and improve conversational data analysis agents using custom datasets and domain-specific questions built for real production environments.

Organizations across IoT, cybersecurity, and e-commerce are moving beyond static dashboards to conversational data agents. These systems promise faster access to insight—but in production, many still fail on the questions that matter most.

That creates a critical business risk: **can your team trust the agent's answers enough to act on them?**

Why current evaluations fall short

Most data agents perform well on simple benchmark tasks, but those tests rarely reflect the complexity of enterprise environments.

- **Generic datasets**
Public benchmarks do not capture the patterns, anomalies, and semantics found in real business data.
- **Low-value queries**
Most evaluations focus on simple questions like averages and counts—not the time-based, incident-driven reasoning required in production.
- **Misleading benchmark results**
Agents that look strong in demos often break on domain-specific questions, creating a reliability bottleneck for deployment.

In internal testing, agents that scored 73% on standard benchmark-style tasks fell to 10–20% accuracy on domain-specific, incident-focused queries.



The Solution: Rockfish AgentFuel

Rockfish AgentFuel is a customizable evaluation framework for time-series data agents. It helps teams move beyond generic testing and measure performance against the data patterns and questions that actually matter in their environment.

What makes AgentFuel different

- **Custom evaluation data**
Generate realistic datasets from your schema or use privacy-preserving synthetic generation based on your data characteristics.
- **Data-aligned question-answer pairs**
Automatically create question-answer pairs linked directly to the generated data, including scenarios involving timing, event sequences, anomalies, and behavioral shifts.
- **Failure-mode visibility**
Identify issues such as schema confusion, weak data exploration, and reasoning breakdowns before agents reach production.

Why teams use AgentFuel

- **Launch faster** with custom benchmarks tailored to your environment
- **Improve reliability** by testing against realistic, domain-specific scenarios
- **Optimize continuously** using evals to improve prompts, workflows, and agent performance
- **Reduce risk** by validating agent behavior before production rollout

Early evidence shows that integrating AgentFuel into an optimization loop can improve agent accuracy by 17–25%.

Don't rely on benchmark scores alone. Evaluate your data agents on the scenarios that matter—and **deploy with confidence using Rockfish AgentFuel**