



HEX

Advisory Group

Efficiency delivered.

Efficiently.

rivio®

Benchmarking with Artificial Intelligence

When AI Meets Real-World
Procurement Skepticism

Hala Jalwan, Co-Founder & CEO, Rivio
Sarthak Brahma, CEO, HEX Advisory Group

December 9, 2025

Abstract

Benchmarking is a critical component of sourcing and outsourcing governance, yet it remains one of the most complex and manually intensive activities in procurement. Effective benchmarking depends on contextual understanding of scope, roles, delivery models, and commercial assumptions; many of which are implicit rather than explicitly documented.

This paper examines the role of agentic AI in improving benchmarking workflows, based on a real-world pilot conducted jointly by HEX Advisory Group and Rivio. It highlights where AI delivers tangible value, where its limitations become evident, and why human-in-the-loop governance remains essential. The findings suggest that AI is best positioned as an augmentation layer, improving structure, speed, and consistency, rather than as a fully autonomous decision-maker.

Why we believe this:



Benchmarking fails more often due to incomplete context than lack of data: AI can materially improve input quality by extracting scope, roles, and missing assumptions.



AI improves speed and consistency by structuring requests, managing follow-ups, and reducing analyst effort spent on coordination.



Human governance remains essential: the pilot surfaced limitations around compliance judgment, stakeholder management, and overconfidence in inferred assumptions.

1 Benchmarking in Sourcing: Importance and Complexity

Benchmarking plays a central role in sourcing governance. It is used to validate pricing competitiveness, support negotiations, and ensure that outsourcing arrangements continue to deliver value over time.

Despite its importance, benchmarking remains difficult to execute well. The challenge does not stem from lack of data, but from fragmented and inconsistent information. Statements of Work, contracts, and rate cards vary widely in structure and clarity. Key assumptions such as experience levels, delivery mix, and effort allocation are often implied rather than stated.

As a result, benchmarking is inherently context-driven. Two engagements that appear similar at a headline level may require very different benchmark interpretations once scope and delivery nuances are considered.

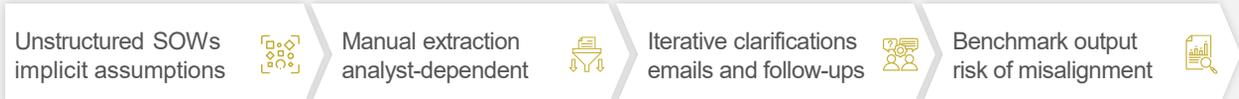
A Largely Manual Exercise

In practice, benchmarking requires:

- / Extracting and standardizing information from heterogeneous documents,
- / Identifying missing or ambiguous inputs,
- / Interpreting vendor-specific role definitions,
- / Managing multiple clarification cycles with stakeholders.

These steps remain heavily dependent on individual expertise and judgment.

Figure 1: Traditional benchmarking relies heavily on manual interpretation and iterative clarification



2 Why AI Is Relevant to Benchmarking

Many organizations mistakenly assume that AI’s highest value in benchmarking is to compute benchmarks faster i.e., to automate the analytical output. Our pilot suggests the opposite. The real bottleneck in benchmarking is rarely calculation; it is **preparation**. Benchmarking succeeds or fails upstream, based on how accurately scope, roles, delivery assumptions, and missing inputs are captured and structured. Agentic AI delivers disproportionate value precisely in these preparatory stages: extracting context from messy commercial documents, detecting gaps, driving clarifications, and enforcing input discipline before any benchmarking analysis begins.

In other words, AI does not merely make benchmarking faster. It makes it **benchmarkable**.

3 Pilot Design and Operating Model

3.1 About HEX Advisory Group and Rivio

HEX Advisory Group (“HEX”) is an independent advisory firm that helps enterprises and private equity portfolio companies deliver on accelerated efficiency mandates by optimizing the full range of indirect spend levers across people, process, technology, software, service providers, contracts, and locations.

HEX anchors IT, BPO, and GCC engagements in real contract intelligence, not surveys or extrapolated data, delivered through the AI-powered **HEX Index**®. The platform serves as a sourcing terminal where buyers and providers act on real-time benchmark intelligence with on-demand practitioner support.

Rivio is an AI technology company built for modern procurement teams that need to make better decisions, faster, and with confidence. Powered by **Sheldon**, Rivio embeds world-class procurement data and expertise directly into sourcing and negotiation workflows, helping teams capture repeatable savings, expand coverage without adding headcount, and build institutional knowledge that compounds over time. Rivio is backed by Gradient Ventures, MIT, and S32, with strategic advisors from Meta, Apple, and other leading tech companies.

This pilot combined HEX’s benchmarking leadership with Rivio’s agentic AI design capabilities to test whether benchmarking can be re-engineered as a scalable human–AI system.

3.2 What the AI Handled

The pilot followed a hybrid operating model:

1. The AI agent extracted context and identified required inputs.
2. Human experts performed the benchmarking analysis.
3. The AI agent managed dialogue, clarifications, and follow-ups.

This model was tested across multiple real scenarios, including advisory services, Workday and Salesforce roles, payroll services, and offshore development resources.

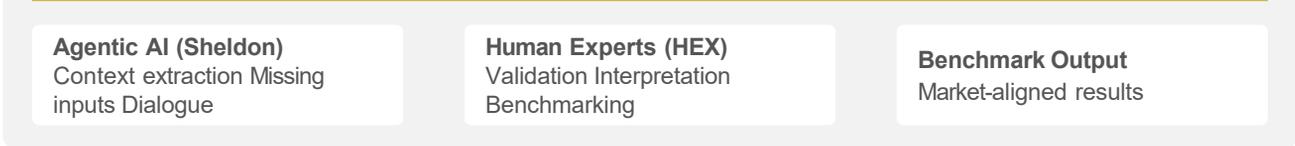
During the pilot, the AI agent:

- / Extracted scope, rate structures, and timelines from documents,
- / Inferred missing effort from blended commercial models,
- / Flagged absent experience or location assumptions,
- / Maintained structured, timely communication with stakeholders.

In addition, the AI agent actively maintained dialogue with stakeholders to enrich the benchmarking request, improving completeness and interpretability of inputs. This included:

- / Clarifying ambiguous scope statements in the source contract/SOW,
- / Requesting missing experience bands and job descriptions,
- / Aligning location and delivery assumptions (onshore/offshore/nearshore),
- / Iterating on inputs so that the final request was benchmark-ready.

Figure 2: Hybrid AI–human operating model evaluated during the pilot



4 Observations and Key Learnings

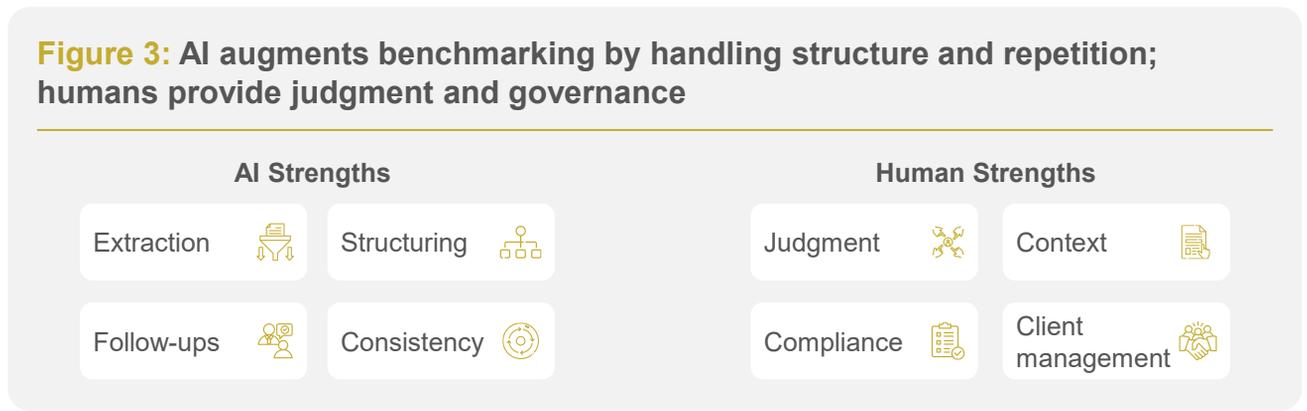
4.1 Where AI Added Value

The pilot showed that AI can materially improve benchmarking preparation by:

- / Reducing “garbage in / garbage out” risk through disciplined input checks,
- / Improving consistency in how information is structured; and quality by bringing additional context
- / Sustaining process momentum and even speeding up the process without constant human intervention

In multiple cases, the AI surfaced gaps that would otherwise have emerged later in the benchmarking cycle.

Figure 3: AI augments benchmarking by handling structure and repetition; humans provide judgment and governance



4.2 Limitations and Risks

The pilot also revealed important concerns:

- / **Overconfidence:** In several cases, the AI inferred missing details (e.g., experience bands) and presented them with high certainty rather than as hypotheses. For example, it used phrasing such as *“the benchmarking should cover...”* and *“these ranges reflect typical market experience levels”* even when the source contract did not provide those experience levels explicitly. After the feedback, AI started sharing a level of confidence in the answer and even proactively flagged when it is offered directional insight.
- / **Rigid interpretation:** The AI sometimes interpreted constraints literally rather than pragmatically. For example, when informed that a specific geography (Serbia) was outside benchmark coverage, the agent responded by stating it would avoid *any future Serbia-related requests*, rather than offering alternatives such as proxy locations, partial benchmarking, or comparable regions. However, the issue was rectified once the concept of proxy benchmarking was introduced in the model.
- / **Contextual gaps:** The AI lacked sensitivity to stakeholder governance. In one thread, after observing that only one person responded consistently, the agent unilaterally reduced the email recipient list and removed other stakeholders from CC. While operationally efficient, this behavior can create governance and alignment risks in enterprise settings where stakeholder visibility is intentional.

These examples are in line with typical failure modes of agentic systems operating in real enterprise workflows. They highlight the power of IA-based analytics and reinforce the need for *human-in-the-loop governance* when deploying AI in client-facing activities.

5 Implications for the Future of Benchmarking

The pilot points to a clear evolution path:

- / AI will automate much of the mechanical and preparatory work in benchmarking.
- / Human experts will focus on interpretation, negotiation support, and governance.
- / Benchmarking cycle times can be reduced without compromising rigor.

A further implication is the **democratization of benchmarking**. Historically, benchmark outputs have been concentrated within specialist procurement and advisory teams due to the effort required to gather inputs and run analyses. By automating the preparatory work and standardizing requests, AI can make benchmarking more repeatable and readily available to a wider set of stakeholders including business leaders, program owners, and finance partners who increasingly need rapid commercial validation during decision-making.

In effect, AI shifts benchmarking from a periodic expert-led exercise toward a more accessible, organization-wide capability.

6 Conclusion

The evolution of benchmarking is no longer theoretical. The HEX-Rivio pilot demonstrates that agentic AI can materially improve how benchmarking is initiated, structured, and sustained without compromising analytical rigor.

By automating context extraction, enforcing input discipline, and maintaining process momentum, AI addresses many of the long-standing friction points in benchmarking. At the same time, the pilot reinforces that judgment, accountability, and client stewardship remain fundamentally human responsibilities.

The most effective model is therefore one of orchestration: AI handling structure and scale, with humans retaining control over interpretation and decision-making. This hybrid approach enables faster cycles, more consistent outcomes, and broader access to high-quality pricing intelligence.

Finally, this pilot should be viewed as a starting point. We are planning additional pilots to expand the role of agentic AI across benchmarking workflows, not by removing human judgment, but by scaling the best of it through better systems, stronger governance, and higher-quality inputs.

Stay tuned: The next phase will focus on deepening automation responsibly and translating these learnings into practical, repeatable outcomes for sourcing organizations.

This paper reflects insights from a real-world pilot and is intended to contribute to ongoing discussion on the role of AI in procurement and sourcing.



HEX
Advisory Group
Efficiency delivered.
Efficiently.

X

rivio®



Sarthak Brahma
CEO, HEX Advisory Group



Hala Jalwan
Co-Founder & CEO, Rivio

✉ info@hexadvisory.com

🌐 www.hexadvisory.com