Token volume isn't the metric. Interaction quality is.

A short brief for technology leaders evaluating AI engineering productivity.

May 11, 2026 · TVCRpro · USPTO 64/045,951

What this brief is

One page on a problem we believe enterprise IT leaders are starting to articulate publicly, and a methodology already filed and built that addresses it.

This is not a sales pitch. It is a conversation prompt.

Tokens are the fuel of AI. TVCR is the mileage.

Every AI interaction burns tokens. A fleet manager who spent thousands on fuel without measuring miles-per-gallon would not call the spend an investment. They would call it consumption. The same standard applies to enterprise AI spend.

The problem, in the words of a CIO who runs 12,000 engineers

In an interview published May 8, 2026, Marco Argenti, Chief Information Officer of Goldman Sachs, described the metric his team uses to evaluate AI productivity:[1]

> "If you have a lot of token usage and output does not move, then at that point, it probably means that you are still in the experimentation phase. We identified a threshold — below it there was no change in the output metrics, but once we surpassed that threshold, productivity started moving."

The Goldman framing is a fuel-economy observation: below a per-interaction threshold, the engine is burning tokens without converting them into outcomes; above it, the conversion ratio jumps. TVCRpro is the framework that measures that ratio for a single interaction at a time, before a quarterly retrospective can confirm whether the spend was investment or consumption.

He went on to explain why that threshold exists:

> "People were going back and forth with the AI on the planning itself — using tokens to create implementation plans and business requirement documents before getting into coding. That preparatory work does not immediately yield coding output, because it happens before developers start writing code."

The CIO's read: token volume is a lagging signal. The leading signal is the quality of the planning conversation that precedes any code being written.

What is actually being measured today

Most AI productivity dashboards today report two things:

1. **Adoption** — how many seats are using the tool.
2. **Volume** — how many tokens are being consumed.

Neither tells a leader whether a given conversation between a human and an AI was a good one. Neither tells a developer how to have a better conversation tomorrow than they had today. Neither gives procurement a defensible answer to *what did this spend buy us*.

The Goldman finding is the right shape of the problem. The missing piece is a per-interaction measurement that turns the planning conversation itself into a scored, auditable artifact.

What TVCRpro is

TVCRpro is a methodology and scoring system for evaluating the quality of a single human-AI interaction. It scores the prompt, the response, and the relationship between them across eleven dimensions, drawn from a published rubric grounded in patent filing 64/045,951 (USPTO, April 21, 2026, Confirmation 7707).[2]

The scoring system runs in two tiers:

- A **Light** tier, public and unauthenticated, that returns a composite score plus three dimension highlights. Anyone can try it at tvcrpro.com/analyzer.
- A **Full** tier, license-gated, that returns the full eleven-dimension breakdown plus per-dimension coaching language and token-normalization metadata.

The methodology, dimension structure, and scoring architecture are protected by patent application 64/045,951. Production weights, anchors, and normalization constants are retained as trade secret.

Why this matters for the CIO problem

Three points map directly onto what Argenti described:

1. **It scores the planning conversation.** The Behavior, Vetting, and Specification dimensions explicitly reward grounding, clarification depth, and provenance — the work a developer does *before* code is generated. This is the activity Goldman has detected as predictive of throughput but cannot currently measure per interaction.

2. **It is per-interaction, not per-person.** The score is attached to the conversation, not the user. A leader can roll up team-level quality trends without an individual surveillance dashboard. This is the same posture Argenti described — team velocity over personal tracking.

3. **It produces an auditable record.** Each scored interaction can be retained in the licensee's environment for retrospective review. This addresses the procurement question of what token spend bought, because it gives an answer per interaction rather than per quarter.

What we are not claiming

We are pre-pilot. There are no production deployments yet. Synthetic validation underpins the published methodology and that is disclosed in the public Limitations document. Calibration to organization-specific anchors is a deliberate Phase 2 activity that follows licensing, not a precondition.

Per the patent, TVCR scores are not designed to be used as the sole basis for consequential employment decisions until Phase 3 deployment status is reached within the licensee's environment.

What a conversation with us looks like

A first call covers fit, target use case, and whether the rubric matches the developer interactions you actually want measured. A mutual NDA is the prerequisite to any methodology detail beyond what is on the public site. Pilot scoping is by application. Public pricing is not posted at this stage.

Reading list

- Methodology, public: tvcrpro.com/methodology
- Limitations, public: tvcrpro.com/limitations
- Licensing pathway and tiers: tvcrpro.com/licensing
- Patent filing: USPTO 64/045,951, filed April 21, 2026

Contact

Licensing: licensing@tvcrpro.com · NDA: nda@tvcrpro.com · Pilot scoping: pilot@tvcrpro.com

— Atomic Armstrong, TVCRpro

Sources

[1] Reed Alexander, "Forget tracking AI use. Goldman's tech boss cares about this instead." *Business Insider*, May 8, 2026.

[2] United States Patent and Trademark Office, application 64/045,951, filed April 21, 2026, Confirmation 7707. Filing receipt available on request under NDA.