

Stop Asking How Big the Model Is. **Start Asking If It Finished.**

Agentic AI just went commercial. The benchmark conversation is fading — traceability is the real currency now.

SWIPE →

Three things traceability now requires

01 See every step the agent takes

02 Give failure a boundary, not just a probability

03 Price by cost per completed task

From Demo to Billed.

Agentic AI has quietly moved into paid, live production.

Insurance teams handling multi-step claims intake. Engineering teams running autonomous debugging on production codebases — not an experiment, something they're billed for every day.

Nice-to-Know.

That's all a benchmark score is now, if you're running critical infrastructure.

What actually matters: did the task get done, can you prove it got done, and if it didn't, did the system fail safely and visibly.

Cost Per Completed Task.

Not tokens, not seats. A simpler, cheaper model that reliably nails a narrow task can beat a flashier frontier model that costs more per completed job.



**A system that can't tell you why a task failed
can't be improved, only replaced.**

THE CENTRAL ARGUMENT

Benchmark era vs. traceability era

BENCHMARK ERA

- How big is the model
- Where it sits on the leaderboard
- Clean, controlled test performance

TRACEABILITY ERA

- Full action reconstruction, step by step
- Controlled failure boundaries
- Auditable task-completion data

Traceability Can't Be Bolted On.

For PUB, LTA, and every IMDA- or MAS-governed deployment, audit logging, escalation thresholds, and outcome verification have to be part of the architecture from day one — not patched in once a regulator asks for it.

Can Your Vendor Show You the Logs?

If a vendor can't produce auditable completion data for a comparable deployment, that's a red flag — no matter how good their model looks on a benchmark chart.

Full piece:

terencekok.com/blog/agentik-ai-commercial-singapore-traceability

Prove the Task Got Done. Or It Didn't.

IMDA's updated Model AI Governance Framework already makes this non-negotiable: auditability, defined failure boundaries, and human oversight for every agentic deployment.



Terence Kok · Enterprise AI Strategist

SAVE · COMMENT · FOLLOW