GOVERNANCE VERIFICATION · MEASUREMENT

GQM as Governance Verification: The Measurement Gap in NIST AI RMF

If you cannot tell leadership whether governance is working, it is not working. That is not a rhetorical point. It is a measurement problem.

Forty years ago, Victor Basili and colleagues at NASA developed Goal-Question-Metric (GQM) to address a problem that should sound familiar. Organizations said their software processes were high-quality. They had no rigorous way to verify it. The framework was simple in principle and demanding in practice: start with the goal, derive the questions that would tell you whether the goal is being achieved, then identify the specific metrics that would answer those questions.

Applied to AI governance, it looks like this:

- **Goal:** Ensure human oversight in agentic AI deployments is substantive, not nominal.
- **Questions:** At what decision points does human review occur? What information does the reviewer have? What proportion of AI-recommended actions are reviewed versus auto-approved? How often does human review change the outcome?
- **Metrics:** Intervention rate by decision category. Review duration relative to action complexity. Outcome divergence rate between AI recommendation and human decision.

This is not complicated, but it is almost never done. The NIST AI RMF GOVERN function calls for accountability and organizational roles. The MANAGE function calls for risk treatment. Neither specifies the measurement infrastructure that would tell you whether either is having a positive effect.

When an AI-assisted decision goes wrong, the investigation almost always finds the same thing: a human was technically in the loop, but the conditions for meaningful review were never actually present. GQM provides the discipline to stop treating oversight as a process checkbox and start treating it as a verifiable organizational capability.

thinkcapital.org · thinkcapital.substack.com/subscribe · GIAG Stream Two

