AI MEASUREMENT · FUNCTIONAL SIZING

Can We Size AI Systems? Adapting Functional Measurement for Non-Deterministic Software

There is a principle that has held up across every era of software development: you cannot manage what you cannot measure (Tom DeMarco, 1982). Capers Jones spent four decades building the empirical case for it. The functional sizing community gave practitioners the tools to act on it. The principle has not changed. The technology has.

Today, government agencies are signing contracts for AI systems without the measurement frameworks needed to specify what they are buying, assess what was delivered, or benchmark whether the investment made sense. This is not a new observation — it is an unsolved one. The consequences of leaving it unsolved get more expensive with every budget cycle.

Traditional functional sizing methods were built for deterministic software. IFPUG Function Points, COSMIC, and Mark II frameworks measure what a system does, independent of how it is built. That independence made them powerful. When identical inputs reliably produce identical outputs, you can count the functions, size the scope, and build a procurement baseline that holds.

AI systems do not behave this way. Model training is not a transaction. Prompt engineering does not fit the entry/exit/inquiry taxonomy. Emergent behaviors are not specified in functional requirements. Non-deterministic outputs mean the sizing target shifts even as you measure it. The foundational assumption — that a system's functional scope can be enumerated, counted, and compared — breaks down before the first data element type is defined.

The management need has not gone away. It has grown. How does a contracting officer write an objective RFP for AI capabilities? How do vendors compete when bids vary 300 percent for functionally similar scope? How does a program manager establish a cost baseline when no historical data exists for the system type?

A hybrid approach may restore the discipline: IFPUG for the external transaction boundary, a COSMIC-derived layer for internal AI processing complexity, and governance weight factors tied to NIST AI RMF requirements. The research question is whether these components can produce consistent, practitioner-reliable estimates across AI system types.

thinkcapital.org/research.html · Government IT and AI Governance Initiative