From Quantifying Communication to Orchestrating It: Talents × HRO Principles as an Immutable Substrate for AI-Augmented Systems Engineering

Alexander I. N. Derkatsch

Independent Researcher · bigBespoke LLC · Huntsville, Alabama, USA

ai.derkatsch@gmail.com · alexanderderkatsch.com · github.com/DrRainbows

AI4SE & SE4AI Research and Application Workshop 2026 — Extended Abstract — Paper Presentation Track

Primary Research Area: (AI4SE) Agentic AI and workflow integration

Secondary Research Area: (AI4SE) AI and cognitive assistants to support tasks from inference, decision-support and requirements trades to project management

ABSTRACT

Systems engineering lacks a comprehensive theory of how a designer comes to understand a system deeply enough to control its emergent properties. In practice, the engineer reaches that understanding through saturated sensory contact with the system — its smell, its sound, its grease, its bite — until the system's regularities register non-linguistically. AI agents are now capable of performing this becoming-the-system work at boundaries the human cannot occupy: the scale of large organizations, the duration of multi-year programs, the partitions of classified data, and many places at once. We propose a substrate for that capability composed of two immutable axes: **thirty-five thinking talents**, immutable per individual in the sense that they are the underlying tendencies from which a person's emergent state-driven personality is generated, and the **five Weick–Sutcliffe High Reliability Organization principles**, immutable behavioral properties that distinguish high-reliability organizations from the rest. Their cross-product yields a **175-cell immutable pattern matrix**. We argue that this is precisely the cardinality required to control a complex system: any complex system is controlled by a small number of underlying parameters — typically two or three, rarely six or seven, and never twenty-two — and the talents-and-HRO pair are the two controlling parameters at the people-organization layer. Personality is not in this matrix because it is, on this account, a parameter-twenty-two-class emergent surface property, not a controlling variable. Sensory or cognitive ranking (kinesthetic, auditory, visual, or any digital-twin sensor-fusion split) is an aspect dimension applied to read the cells — determinate-under-context — not a multiplier of them. We synthesize five production AI systems built by the author from 2024–2026 against the existing literature, identify gaps where the substrate is empirically ahead, and outline a three-phase validation strategy. The framework offers a path from *quantifying* communication in high-reliability organizations to *orchestrating* it through an AI substrate whose recommendations are traceable to immutable structural axes rather than to model intuition alone.

1. Objectives and Motivation

The workshop theme — *"Speed and Innovation through AI4SE & SE4AI"* — asks how AI-augmented engineering loops can deliver mission capability at higher cadence without sacrificing the discipline that makes high-reliability operation possible. We address that question by locating a missing piece of systems engineering theory and supplying it.

The missing piece is the path from observation to control of emergence. To make a system *do* anything one must control its emergent properties; to control its emergent properties one must understand the system deeply enough to predict them; and the deep understanding required is, in practice, gained only through total immersion — sensory contact saturated enough that the engineer begins to hear the system speak non-linguistically. There is no published comprehensive theory of how an engineer comes to do that. There is no published canonical method for offloading the becoming-the-system requirement to a non-human substrate when the system exceeds human scale, duration, or admissibility.

This work proposes both: a structural theory grounded in two immutable axes, and an AI substrate that performs the immersion at the boundaries the human cannot occupy. Five production AI systems shipped by the author over 2024–2026 are surveyed as the empirical triangulation, with literature synthesis identifying where they extend existing work.

2. Background: The 2021–2022 Frameworks

Two prior peer-reviewed works establish the inputs to the proposed substrate. The first, presented at the AIAA SciTech Forum in 2022, formalizes the organizational hierarchy as `person ⊂ team ⊂ department ⊂ {organization | company}` with a corresponding decomposition of communication artifacts; its central operational proposition is that as communication across the hierarchy increases, the impact of unexpected events decreases, conditioned on adherence to the five Weick–Sutcliffe HRO principles — *preoccupation with failure, reluctance to simplify, sensitivity to operations, commitment to resilience, deference to expertise*. The mediating capability the framework implicitly required — an agent that could interpret communication, classify artifacts against HRO principles, respect individual cognitive priors, and surface emergent patterns at organizational scale — was not physically realizable at the time of publication.

The second, presented at the INCOSE International Symposium in 2021, couples this hierarchy to a five-step sensemaking practice and to the Drivers-of-Thinking taxonomy of approximately thirty-five named *thinking talents* distributed across four cognitive quadrants. After the 2021 presentation, the longtime INCOSE Fellow Dorothy McKinney remarked from the audience that the work was likely "five years ahead of where the field would catch up." The intervening years have produced multi-agent large-language-model systems — the medium of this paragraph's composition — which constitute, exactly five years on, the mediating substrate the 2021 framework named without yet having available.

3. The Immutable Substrate: Two Controlling Parameters, Not Twenty-Two

The proposed substrate is the cross-product of two immutable axes:

$$\mathrm{M} = \mathrm{HRO} \times \mathrm{T}, \quad |\mathrm{M}| = 5 \times 35 = 175$$

where HRO is the five Weick–Sutcliffe principles and T is the thinking-talent catalogue. The cardinality of the substrate — two axes — is itself the central claim. Any complex system is controlled by a small number of underlying parameters: typically two or three, rarely six or seven; beyond roughly ten parameters a system is no longer complex in the controllable sense but chaotic. The talents-and-HRO pair are the two controlling parameters at the people-organization layer of an engineering enterprise. They sit at parameter positions one and two of the people-organization control problem. Personality is not in this matrix because it is — as the Fleeson whole-trait literature confirms — a parameter at position twenty-two or beyond: a fully state-driven, emergent surface property, indeterminate and uncontrollable as a control variable even if observable as a descriptive one.

Talents are immutable in a constructive sense. A child who will stack blocks for hours without burning out, and grow visibly cranky when prevented, has exhibited a well-formed engineering talent before any training is possible. The same underlying talent, given thirty years of experience, will surface as an engineering practice that looks nothing like block-stacking, but the energetic signature is the same: the activity energizes rather than depletes. We adopt without apology the position that personality itself is state-driven and emergent — Fleeson's density-distribution finding that 50–75% of personality-state variance is goal-driven is entirely consistent with this account, because the substrate that the variable state acts upon is the talent profile, not the personality. Coupling experience to expression is not naïve; it is exactly the relationship the model predicts. A person whose underlying talent and current activity are aligned accumulates competence at a much faster rate than a person who lacks the matching talent.

HRO principles are immutable in an organizational sense in the inverse direction: an organization either behaviorally manifests preoccupation with failure, reluctance to simplify, sensitivity to operations, commitment to resilience, and deference to expertise — or it fails to. The principles themselves are stable descriptors of the high-reliability behavioral signature, validated empirically across twenty-five years of Weick, Sutcliffe, Roberts, Schulman, and the Berkeley HRO project.

> *The two-way immutability — talent at the individual level, HRO principles at the organizational level — is what makes this substrate a candidate for the first comprehensive theory of systems engineering. The cells do not depend on state, on personality, or on mood. They depend on what the contributor is innately, and on what the organization behaviorally is.*

A sensory or cognitive ranking — kinesthetic, auditory, visual, or any sensor-fusion split a digital twin admits — is an *aspect dimension* applied to read each cell, not a multiplier of the cell space. The aspect ordering is itself immutable in its underlying hardware preferences, but its expression is *determinate-under-context* rather than stochastic: a contributor whose primary modality is visual at 7 AM may be auditory by evening after caffeine and conversation. This determinacy is what distinguishes the aspect dimension from emergent personality. We make no claim that matching instructional content to VAK modality improves learning outcomes — the claim the Pashler et al. (2008) and Coffield et al. (2004) reviews refute. We claim only that sensory ranking is one available aspect projection through which the immutable 175-cell substrate is read; any other sensor-fusion split could serve, and the choice is empirical.

Engineered systems emerge from the elegance of the people-organization substrate that produces them. The matrix is therefore not an HR instrument. It is the upstream control parameter for whatever the organization intends to build, the substrate whose elegant composition yields elegant emergent systems regardless of target complexity.

4. The AI Agent as Immersion Substitute

The proposed agent architecture is functionally a digital-twin reader for the matrix. It ingests artifacts at the boundary of a system — meeting transcripts, requirements repositories, change requests, hazard logs, telemetry streams, design-review records — interprets them against the 175 cells, and surfaces which cells are currently active for which contributors and which cells should be active but are not. The agent performs the becoming: saturated sensory immersion at boundaries the human cannot occupy, then a human-legible report on what is active and where the organization is gapped against its own HRO-principle topology. Philosophically, this is the extended-mind thesis of Clark and Chalmers operationalized: the AI is functionally coupled to the engineer's cognitive process, extending perception and pattern recognition into domains the engineer cannot physically occupy, while preserving the engineer's tacit Polanyian knowing as the interpretive center.

Three architectural commitments distinguish this agent from generic LLM assistants. *First*, the immersion is **deterministic at its timing layer**: a monitor — a small, auditable state machine — decides when the LLM is invoked, what subset of state it sees, and what prompt frame it receives. The LLM decides only what to say and how. The phrase *"the monitor decides WHEN, the LLM decides WHAT and HOW"* is the load-bearing architectural rule, and is corroborated independently by Felix-Cuadras et al. (2026) for industrial control. *Second*, **data-sensitivity is a first-class routing primitive**: artifacts marked restricted or classified are routed to an air-gapped local open-weight model; unclassified artifacts route to a higher-capability cloud provider; the same prompt frame and same matrix schema are honored in both cases. *Third*, **the matrix itself is the agent's evaluation rubric**: the agent's outputs are scored against the 175-cell schema for coverage, calibration, and false-activation rate, providing the continuous test-and-evaluation surface that SE4AI work has otherwise had to construct ad hoc.

5. Empirical Triangulation: Project Synthesis Against Existing Literature

Five production AI systems built by the author from 2024–2026 instantiate disjoint components of the proposed architecture. Each is presented by its unique architectural discovery, followed by a synthesis against the closest existing literature and a call where the system is empirically ahead.

T-MINUS — monitor-driven multi-agent mission control. Eight LLM console operators under a Flight Director command loop, executing the rule *"the monitor decides WHEN; the LLM decides WHAT and HOW."* Closest existing work places role-specialized agents under structured handoff protocols (MetaGPT, ICLR-24; ChatDev, ACL-24); enforces agent safety at the token level (AGENT-C, 2025) or post-generation (ShieldAgent, 2025); and diagnoses multi-agent failure (Cemri et al., NeurIPS-25, MAST taxonomy of 14 failure modes). Hybrid deterministic-stochastic patterns are corroborated independently in industrial process control (Felix-Cuadras et al., MDPI AI 2026). T-MINUS is empirically ahead in two places: deployment in a real-time physics-driven safety-critical loop with validated cost bounding (448 LLM calls across an eight-flight Monte Carlo campaign), and **deliberate engineering of emergent failure cascades from physics-grade part models as high-fidelity training signal** — an inversion of the literature's framing of emergent failure as defect (Park UIST-23; Roig 2025).

Universal Interface — sensitivity tag as a routing primitive. Discrete data-classification tag selects between cloud and air-gapped local providers at prompt-frame level, with identical output shape across tiers. Existing work treats sensitivity as a continuous learned signal requiring differential privacy (PRISM, AAAI-26) or secure multi-party computation (PPRoute, 2026). SAGAI-MID (2026) generates adapter code for known service contracts; LLMLoop (ICSME-25) iterates code-against-tests but assumes the spec is given. Universal Interface is ahead in coupling I/O discovery with adversarial test synthesis in a single closed loop and in compositing AST-policy enforcement with subprocess isolation on every generated connector — a combination absent from both Aethelgard's learned capability governor (2026) and Blyth et al.'s static-analysis feedback (2025).

CopApp.ai — Dialog / World / Psyche decoupled agents with skill-graph decay. Three LLM agents each maintain an independent state representation of a single training scenario. Park et al.'s Generative Agents (UIST-23) scales agents *across* NPCs; Li et al.'s Theory-of-Mind benchmark (EMNLP-23) measures belief-state tracking but does not operationalize it as product. CopApp's *three-agents-modeling-one-character* architecture is structurally novel in the surveyed literature. Skill-decay credentialing combines DAS3H-style per-skill forgetting curves (Choffin, EDM-19) with tiered certification in a deployed competency substrate — not present in PACE (2026), Violakis (2025), or any other surveyed power-imbalanced-authority training system. Chen, Tan & Lee (NAACL-25) document precisely the bias risk persona-prompted LLMs amplify under power asymmetry — the risk CopApp's Psyche-agent decoupling is designed to mitigate.

PawTrek — voice-first narrative aligned with Florida v. Harris at generation time. Voice-first capture produces a structured narrative whose phrasing is shaped, at the moment of generation, to satisfy the probable-cause reliability standard from *Florida v. Harris*, 568 U.S. 237 (2013). Field et al. (Interspeech-23) and Srivastava et al. (IEEE SLT-24) treat law-enforcement ASR as transcription or post-hoc analysis; no surveyed paper closes the loop from speech to admissible artifact in a single field interaction. The peer-reviewed academic literature on CMEK architecture patterns for CJIS-class compliance is **effectively absent** — a publishable gap; PawTrek's design-time CMEK with real-attestation App Check has no academic precedent.

FletchGNC + Beyond Proportional Navigation — doctrine-code coupling. A 44-chapter technical reference book authored alongside a working simulator, each chapter validated against the same RK4 / EKF / PN / plant-inversion stack. LearnLM (2025) augments existing finished text; Knuth's literate programming (1984) co-evolves prose and code without physics-in-the-loop validation; ESA's USACDF (Strauch et al., 2018) relegates the explanatory layer to downstream documentation. Among GNC-specific work, Zipfel's CADAC++ (2014) is the open 6-DOF gold standard; Tipan et al. (JGCD 2020) apply NDI without dynamic-pressure scheduling; Strub et al. (JGCD 2018) use classical LPV scheduling without plant inversion. FletchGNC's specific combination — plant inversion with dynamic-pressure-based automatic gain scheduling for low-Reynolds-number tactical projectile aerodynamics — is a genuine open-literature gap.

6. Expected Outcomes and Validation Strategy

The expected operational outcome is a deployable AI assistant whose recommendations span (a) meetings collapsible because no cell activates, (b) artifacts whose cell activation suggests redaction, rework, or redistribution, (c) IPT seat assignments whose talent coverage does not span the cells the active engineering problem requires, and (d) longitudinal trends in HRO-principle adherence mapped to cells where the agent's confidence is lowest.

Validation proceeds through three phases. **Phase 1 — synthetic replay**: the agent ingests anonymized historical artifacts and produces ranked recommendations rated by an HRO subject-matter panel on a calibrated rubric with inter-rater agreement reported. **Phase 2 — shadow deployment**: the agent runs alongside an active team for one quarter, producing recommendations the team may ignore, with team members surveyed on counterfactual decisions. **Phase 3 — limited-trial intervention**: the agent's surface is integrated into existing tooling under explicit organizational opt-in; measured deltas are unexpected-event impact, IPT communication latency, and HRO-principle adherence over two quarters. The validation methodology is designed to address the AHRQ (2024) finding that HRO-implementation evidence is weaker than the theory's intellectual influence would suggest — by producing quantitative outcome data calibrated against an explicit structural model.

7. Relevance to Practice and Workshop Theme

The framework operates squarely in the workshop's speed-without-sacrificing-rigor intersection. It is grounded in a peer-reviewed communication and collaborative-intelligence formalism, made deployable by a now-available agentic substrate, and triangulated by five shipped production systems. Topcu et al. (Systems Engineering, 2025) document precisely the failure pattern this substrate is designed to catch: LLMs generate expert-like SE artifacts on automated metrics yet exhibit premature-definition, unsubstantiated-estimate, and overspecification failure modes — drift-into-failure patterns that HRO-principle-traceable evaluation surfaces detect by construction. The work extends the AI4SE / SE4AI roadmap (McDermott et al., 2020; Pepe et al., 2022) by populating the human side of human-machine co-learning with a specific structural model, and offers the cognitive-talent substrate that the HAIC-MM maturity-assessment work (Ortolano & Gallegos, 2025) measures the surface of without explaining mechanistically. Attendees will leave with the 175-cell schema (permissive license), the monitor-driven pipeline reference, a sensitivity-routed provider router, and a validation playbook.

8. Honest Limitations and Future Work

The matrix cells are not strictly orthogonal (*Reluctance-to-Simplify* × *Strategy* overlaps with *Sensitivity-to-Operations* × *Thinking-Ahead*); a factor-analysis pass should identify the dominant subset. The Drivers-of-Thinking talent taxonomy itself has not been independently psychometrically validated, and the construct-validation work against the Big Five and CliftonStrengths benchmarks is open future research. The parameter-cardinality claim — that talents and HRO principles are precisely the two controlling parameters of the people-organization layer — is a hypothesis derived from complex-systems theory, not an empirical result; the validation strategy in §6 is its first test. Ethical concerns (observation, consent, opt-out) are deployment preconditions; open-source release facilitates scrutiny.