

## Judgment Engineering // Human Judgment Engineering

### The Human Looping Through the System

Judgment Engineering is the deliberate discipline of design, development, and application of human judgment throughout AI-augmented work so people remain capable of applying subject matter expertise, experience, framing problems, establishing intent and criteria, challenging assumptions, evaluating evidence, governing AI-enabled actions, making consequential and grounded decisions, and learning from outcomes.

The discipline is built on a simple premise: human judgment should not exist as a single approval point at the beginning or end of an AI process moving away from the 'sandwich model' of Human-AI engagement. Human judgment and inputs should repeatedly enter, influence, challenge, redirect, authorize, and learn from AI systems, tools, models, and apps wherever judgment matters. This extends the Human-in-the-Loop (HitL) concept into an architecture of recurring human engagement across conversations with large language models, individual agents, and complex agentic systems. Judgment Engineering is the human decision system and decision environment, including checkpoints, feedback loops, constraints, verification, escalation, reflection, productive cognitive friction, that results in an AI output that meets the needs of the user or client all the while the human manager, leader, or orchestrator is learning, growing, developing, and becoming a more powerful contributor for the organization and for their future selves.

Judgment Engineering therefore goes beyond getting better AI outputs. Prompt Engineering improves how humans communicate with AI; Judgment Engineering improves the broader human decision system surrounding AI: problem framing, assumptions, evidence, uncertainty, consequences, verification, reliance, escalation, and accountability. The ultimate goal is appropriate reliance, ethical conduct, and governance: knowing when to accept, modify, reject, investigate, seek Subject Matter Expertise (SME), escalate, or stop the use of AI based on uncertainty, evidence, expertise, reversibility, and consequences.

#### **Governing Principle:**

AI can contribute intelligence, analysis, generation, recommendations, and action.

Humans retain responsibility for purpose, wisdom, authorization, ethics, and consequences.

### **The Judgment Engineering Architecture**

#### **FORGE - How Human Judgment Is Developed**

**F - Frame the Problem.** Develop the experience and SME to define the real problem, decision to be made, partnerships to consider, criteria for success, constraints, stakes, and the relevant decision rights before pursuing a convenient answer.

**O - Open Alternatives.** Resist premature certainty by deliberately exploring different explanations, perspectives, scenarios, approaches, and/or courses of action.

**R - Review Evidence.** Build the habit of distinguishing claims from evidence, testing sources and reasoning, recognizing uncertainty, and determining what deserves confidence and escalation.

**G - Govern the Decision.** Practice making and defending decisions under uncertainty while balancing expertise, values, risk, consequences, accountability, and appropriate AI involvement.

**E - Evolve Through Feedback.** Strengthen judgment through experience, mistakes, reflection, disagreement, feedback, recalibration, and repeated exposure to consequential decisions.

**FORGE is the development mechanism:** Judgment is built through SME, productive friction, deliberate practice, reflection, and experience. This methodology/model explicitly treats practice, challenge, feedback, iteration, and reflection as mechanisms through which stronger judgment develops.

### **LOOP - Where Human Judgment Enters the System**

**L - Locate the Judgment Points.** Identify where human judgment is required because the work involves uncertainty, material consequences, specialized expertise, ethical considerations, competing priorities, governance requirements, decision authority, or meaningful risk.

**O - Orient the System.** Establish the purpose, intent, context, partnerships, criteria, constraints, values, and desired outcome before allowing the AI system to determine the direction of the work.

**O - Oversee and Intervene.** Remain able to question, challenge, correct, constrain, approve, escalate, or stop the system as new information, uncertainty, or consequences emerge.

**P - Process the Feedback.** Examine results and outcomes, capture what was learned, recalibrate assumptions and standards, and feed that learning back into the next iteration of the human-AI system.

**LOOP is the human judgment system architecture:** Humans do not occupy one box in the process; human judgment repeatedly enters wherever it adds value or protects against consequential error.

### **STEER - What Humans Do While Governing AI-Enabled Work**

**S - Set Intent.** Define what the AI tool, system, model, or app is intended to accomplish, why the outcome matters, how success will be measured, and which decisions or controls must remain under human authority and governance.

**T - Test Assumptions.** Surface and challenge the assumptions being made by both the human and the AI before those assumptions become the foundation of a recommendation or action.

**E - Evaluate Evidence.** Examine the quality, relevance, completeness, uncertainty, and competing interpretations of the evidence and data supporting the AI's conclusions.

**E - Exercise Judgment.** Integrate evidence with subject matter expertise (SME), experience, ethics, context, partner impact, and possible consequences to determine what should actually happen.

**R - Recalibrate.** Adjust the decision, system boundaries, instructions, criteria, or level of AI reliance when evidence, environmental conditions, outcomes, or human understanding change.

**STEER is the governance behavior:** AI provides capability and momentum, humans remain responsible for direction.

### **JUDGE - How Humans Evaluate AI-enabled or AI-supported Outputs**

**J - Jurisdiction.** Determine what decision is actually being made, whether AI is appropriate for that decision, who has authority to act, and who remains accountable for the outcome/result.

**U - Unpack the Output.** Separate the AI's claims, assumptions, evidence, interpretations, recommendations, and omissions rather than the single polished response as a trustworthy output.

**D - Demand Evidence and Dissent.** Require supporting evidence, alternative explanations, counterarguments, missing perspectives, possible failure modes, and information that could challenge or disprove the preferred recommendation.

**G - Grade the Reasoning.** Evaluate whether the reasoning is logical, sufficiently supported, contextually appropriate, ethically defensible, proportionate to the risk, and strong enough to influence action.

**E - Exercise Human Judgment.** Decide whether to accept, modify, reject, investigate, seek SME, escalate, or stop the AI-supported course of action and own the decision.

**JUDGE is the evaluation discipline:** AI produces possibilities and recommendations; humans determine what deserves trust, reliance, action, and publication.

### **One Discipline. Four Functions.**

Judgment Engineering is the overarching discipline.

- FORGE explains how judgment capability is developed.
- LOOP determines where humans engage.
- STEER defines how humans govern the system.
- JUDGE provides a method for evaluating AI-supported recommendations and decisions.

Judgment Engineering is the system.

The Four Functions are the elements that create the details and decision points of the system.

FORGE human judgment →

LOOP judgment throughout the system →

STEER AI-enabled work →

JUDGE what the system produces →

FORGE stronger human judgment.

This creates a continuous learning cycle rather than a one-time approval process.

Human agency becomes structural: humans frame, govern, override, escalate, decide, learn, and own consequences throughout the system.

### **The Human Capabilities Behind the System's Architecture**

Judgment Engineering depends on distinctly human and transferable capabilities:

- Critical reasoning
- Metacognition
- Subject-matter expertise
- Experience
- Leadership judgment
- Ethical reasoning
- Communication
- Systems thinking
- AI awareness, literacy, and fluency
- Accountability
- Human agency

These capabilities determine whether humans can do more than operate AI tools. They determine whether people can govern what those tools produce and do.

Better decisions. Stronger humans. Better systems.