Task Class Differentiation Under Confidence Interruption

by Nick Clark | Published March 27, 2026 | PDF

Distinct interruption protocols for terminal, exploratory, and generative tasks, each receiving appropriate handling when confidence drops below threshold.

What It Is

Distinct interruption protocols for terminal, exploratory, and generative tasks, each receiving appropriate handling when confidence drops below threshold.. This mechanism is defined in Chapter 5 of the cognition patent as a structural component of the agent's cognitive architecture, operating through deterministic evaluation rather than heuristic approximation.

Every aspect of this mechanism is specified declaratively in the agent's policy reference, making it auditable, reproducible, and governable without requiring access to the agent's internal decision-making process.

Why It Matters

Without task class differentiation under confidence interruption, execution is either permanently authorized or permanently blocked. Current systems lack a continuous, computed authorization mechanism that adapts to changing conditions in real time. The result is that agents either act when they should pause, or pause when conditions have sufficiently improved to warrant resumption.

This matters most in safety-critical domains: autonomous vehicles, medical systems, financial trading, and any context where acting under insufficient confidence produces harm. A system without confidence governance either acts recklessly or is so conservative that it fails to provide value.

How It Works Structurally

As defined in Chapter 5 of the cognition patent, task class differentiation under confidence interruption operates through a deterministic evaluation function embedded within the agent's cognitive architecture. The function receives structured inputs from the agent's canonical fields and produces outputs that govern subsequent processing stages. Every input, computation step, and output is recorded in the agent's lineage, ensuring complete reproducibility.

The confidence value is recomputed at each evaluation cycle from current inputs. The computation function is deterministic: given identical inputs, it produces identical outputs. Rate-of-change tracking uses first and second derivatives to detect trajectory trends. Hysteretic thresholds with configurable margins prevent oscillation near authorization boundaries.

What It Enables

This mechanism enables agents that pause when conditions warrant it and resume when conditions improve, without external intervention. The result is autonomous systems that are self-regulating: they do not act beyond their assessed competence, and they do not remain idle when competence is restored.

Because this mechanism is policy-governed and deterministic, it can be formally analyzed, audited, and certified. Regulatory compliance is demonstrable through structural analysis rather than solely through empirical testing. Different domains can tune the mechanism's parameters through policy configuration without requiring architectural changes, making the same structural capability applicable to autonomous vehicles, companion AI, therapeutic agents, and enterprise systems.

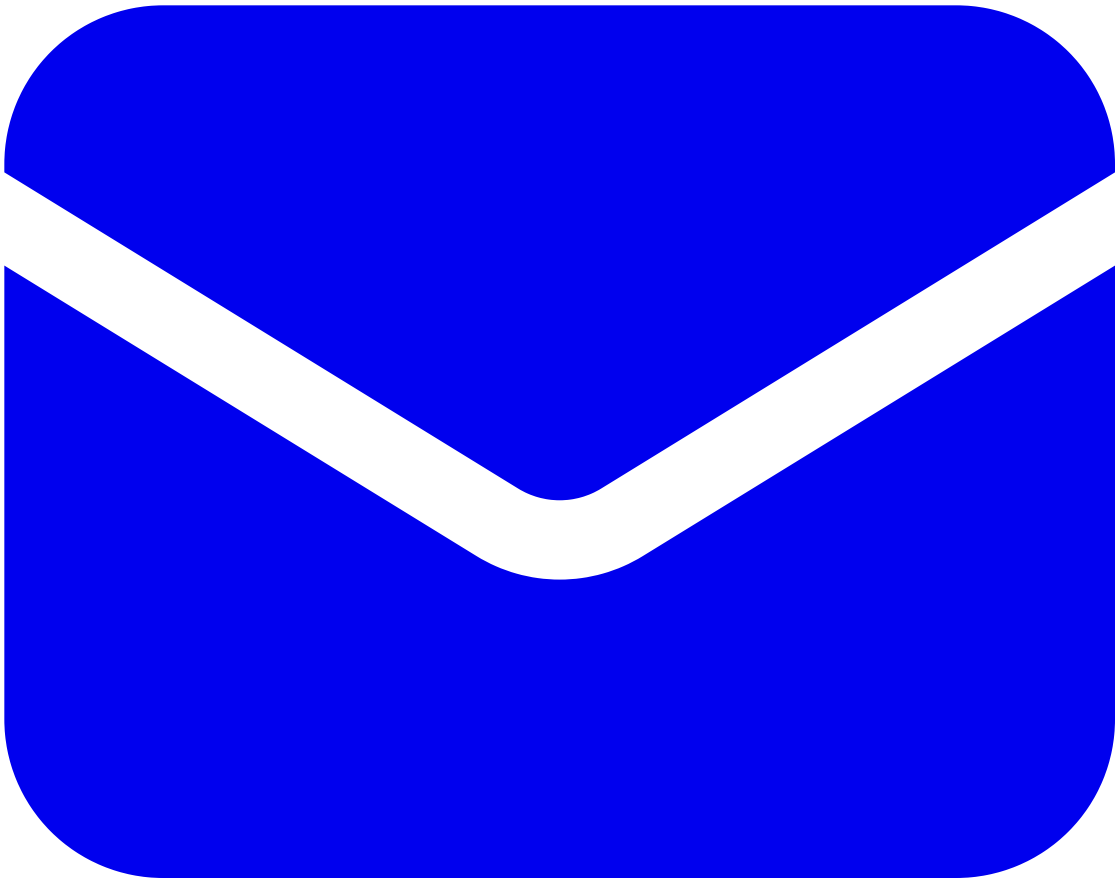
Confidence Governance All 21 steps →

Execution is a revocable permission, not a default.

Primary Technical Disclosure
○ Confidence-Governed Execution: When Agents Pause, Reassess, and Resume Safely
Secondary Technical
○ Execution as Revocable Permission○ Confidence as First-Class Computed State Variable○ Composite Admissibility Evaluator○ Confidence Trajectory Projection○ Non-Executing Cognitive Mode● Task Class Differentiation Under Confidence Interruption○ Confidence-Integrity Feedback Loop○ Differential Rate Alarm Conditions○ Hysteretic Authorization Recovery○ Confidence Computation Function○ Confidence-Driven Inquiry Mode○ Curiosity as Confidence Modulator○ Affect-Modulated Confidence Sensitivity○ Effort Analysis and Path Optimization○ Confidence-Modulated Discovery Traversal○ Biological Signal to Confidence Coupling○ Multi-Agent Confidence Propagation○ Confidence-Governed Embodied Execution○ Deferred Execution and Temporal Reauthorization○ Execution Authorization Recovery○ Confidence Contagion in Delegation○ Confidence History Calibration○ Attention Field
Applications (General)
○ Autonomous Vehicle Execution Safety Through Confidence Gating○ Clinical AI That Pauses When It Should Not Act○ Confidence Governance for Nuclear Operations○ Confidence Governance for Aviation Autopilot Systems○ Confidence Governance for Pharmaceutical Dosing Systems○ Confidence Governance for Bridge Structural Monitoring○ Confidence Governance for Food Safety Inspection○ Confidence Governance for Chemical Plant Operations
Applications (Specific)
○ Agentforce Executes by Default○ Microsoft Copilot Has No Confidence State○ OpenAI Operator Cannot Govern Its Own Execution Authority○ Claude's Safety Has No Computed Confidence Variable○ Gemini's Multimodal Confidence Is Not Computed○ Cohere Command Generates Without Computed Confidence○ AWS Bedrock Guardrails Filter Output Without Governing Confidence○ Azure Content Safety Classifies Harm Without Governing Execution○ Google Vertex AI Safety Filters Without Confidence State○ NVIDIA NeMo Guardrails Constrains Dialogue Without Governing Confidence○ Guardrails AI Validates Output Without Governing Execution Authority○ Lakera Guards Inputs Without Governing System Confidence
Confidence Governance overview →
AQ
deterministic
autonomy

Legal

Last updated: 2026-03-03



-
- nick@qu3ry.net
- 72 28 14 36 01

Invented by Nick Clark | Founding Investors: Devin Wilkie