

Security Implications

The meta-level framework has significant implications for cognitive security, AI safety, and the robustness of belief systems.

Understanding meta-cognitive processing reveals vulnerabilities that traditional security models miss, while also suggesting principled defense strategies.

Cognitive Security Framework

Active Inference's quadrant structure provides a systematic way to analyze cognitive vulnerabilities. Each quadrant represents a potential attack surface with distinct vulnerability profiles and defense requirements.

Attack Surface by Quadrant

| Quadrant | Target | Vulnerability | Impact |
|----------|--------------|-----------------------------|---------------------------|
| Q1 | Sensory data | Observation manipulation | Belief distortion |
| | Meta-data | Quality score falsification | Confidence miscalibration |