Deskilling Traps: The Hidden Long-Term Risks of AI Dependence

A Dynamical Systems Analysis of Supervisory Skill Erosion

BASED ON THE RESEARCH "DESKILLING TRAPS: A DYNAMICAL SYSTEMS MODEL OF SUPERVISORY SKILL EROSION UNDER AI ASSISTANCE"

Immediate productivity gains mask a critical open problem regarding long-term capability

The Gain

The Evidence for Speed

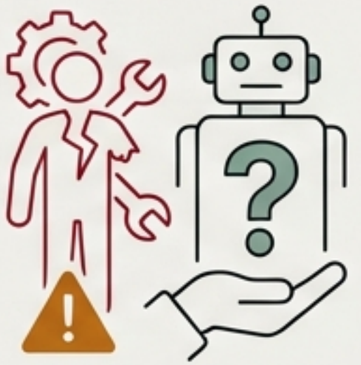
- Software engineers complete tasks faster (Peng et al., 2023)
- Knowledge workers produce higher quality work (Dell'Acqua et al., 2023)
- Medical professionals achieve greater accuracy on routine cases (Brynjolfsson et al., 2023)

The Tension

The Open Problem

It is unclear whether the use of AI assistance might prevent the development of skills necessary to supervise automated tasks.

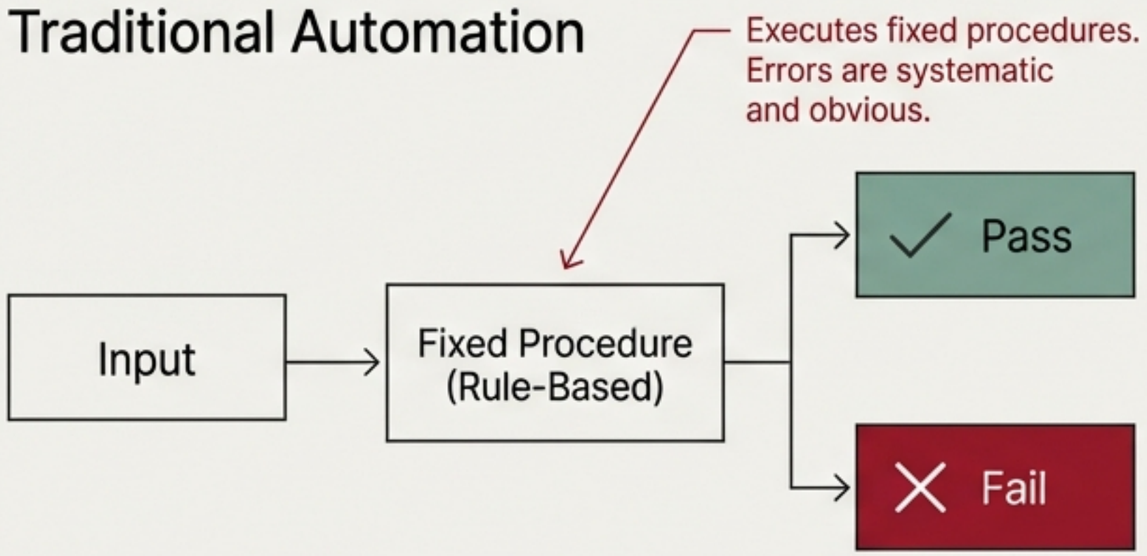
— Shen et al. (2026)



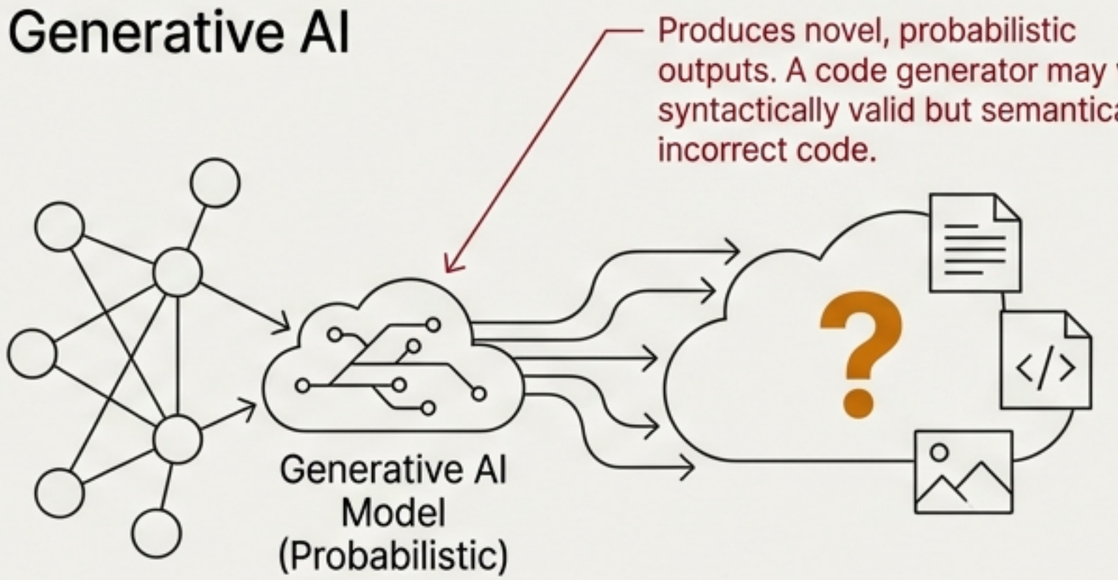
! If we offload the doing, do we lose the ability to check the doing? !

Unlike traditional automation, Generative AI requires judgment to supervise—the exact skill it erodes.

Traditional Automation



Generative AI

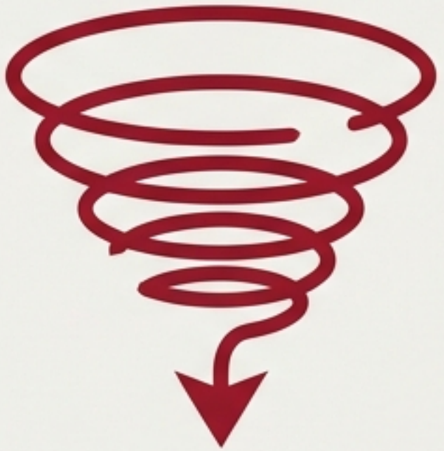


Detecting these subtle errors requires deep domain expertise. By offloading the task, we erode the feedback loop required to maintain that expertise.

A "Deskilling Trap" occurs when a worker loses both competence and the awareness of their incompetence.

Supervisory Skill (s) < 0.3
(Inability to effectively supervise)

Metacognition (m) < 0.3
(Unaware of the inability)



Based on the Dunning-Kruger effect, this creates a “Zone of No Return.” Once a worker enters this state, they cannot self-correct because they no longer recognize that their reliance on the AI is unsafe.

To diagnose this risk, we model the worker as a dynamic system of three interacting variables.

Skill (s)

Evolved through practice, decays with disuse. Modeled via logistic growth vs. exponential decay dynamics.

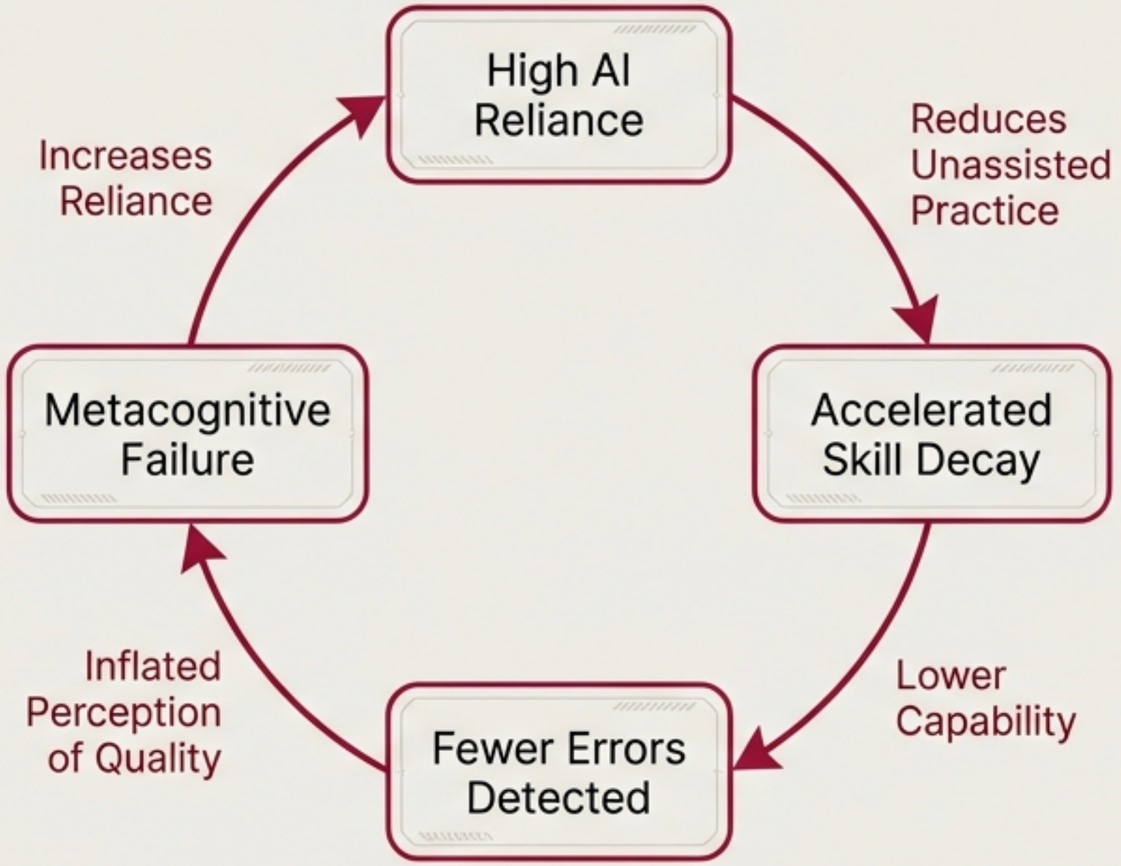
Metacognition (m)

The “calibration” of self-trust. Grows when errors are detected; decays into complacency when AI is highly reliable.

Reliance (r)

How much the worker delegates to AI. This is endogenous—it adapts automatically based on perceived AI quality.

High AI reliance creates a feedback loop that accelerates skill decay and creates false confidence.



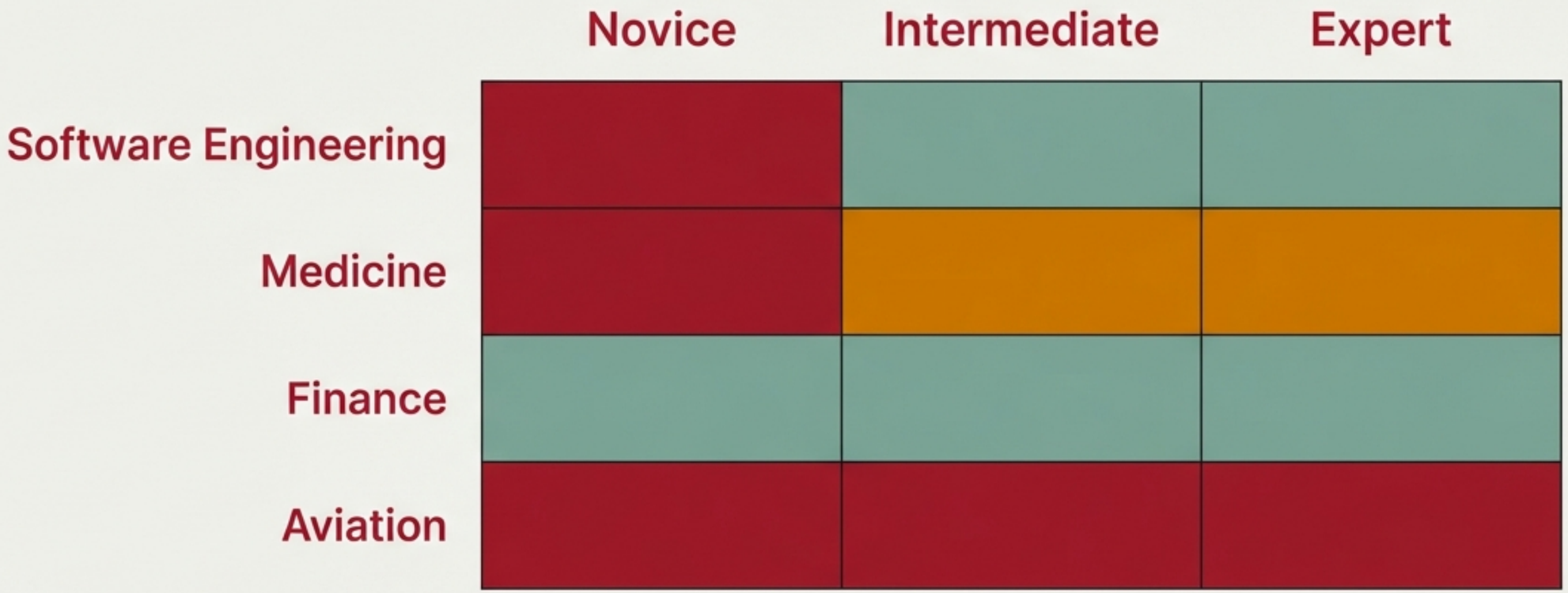
The worker trusts the AI more because they are becoming less capable of seeing its mistakes.

We simulated 200 weeks of work across four distinct professional domains.

Domain	Key Parameters
Software Engineering Crimson Pro	High novelty (0.25), Moderate consequence. Crimson Pro
Medicine Crimson Pro	High consequence (0.90), High reliability. Crimson Pro
Finance Crimson Pro	High novelty (0.30), Delayed feedback. Crimson Pro
Aviation Crimson Pro	Extreme reliability (0.95), Catastrophic error severity. Crimson Pro

Model calibrated using parameters from human factors literature (e.g., Bainbridge, Parasuraman & Riley).

Aviation and Medicine are the most vulnerable domains, while Finance is protected by high novelty.



LEGEND: Red = Trapped (Skill < 0.3), Green = Safe.

In high-reliability domains, experts are not immune to the deskilling trap.

0.95

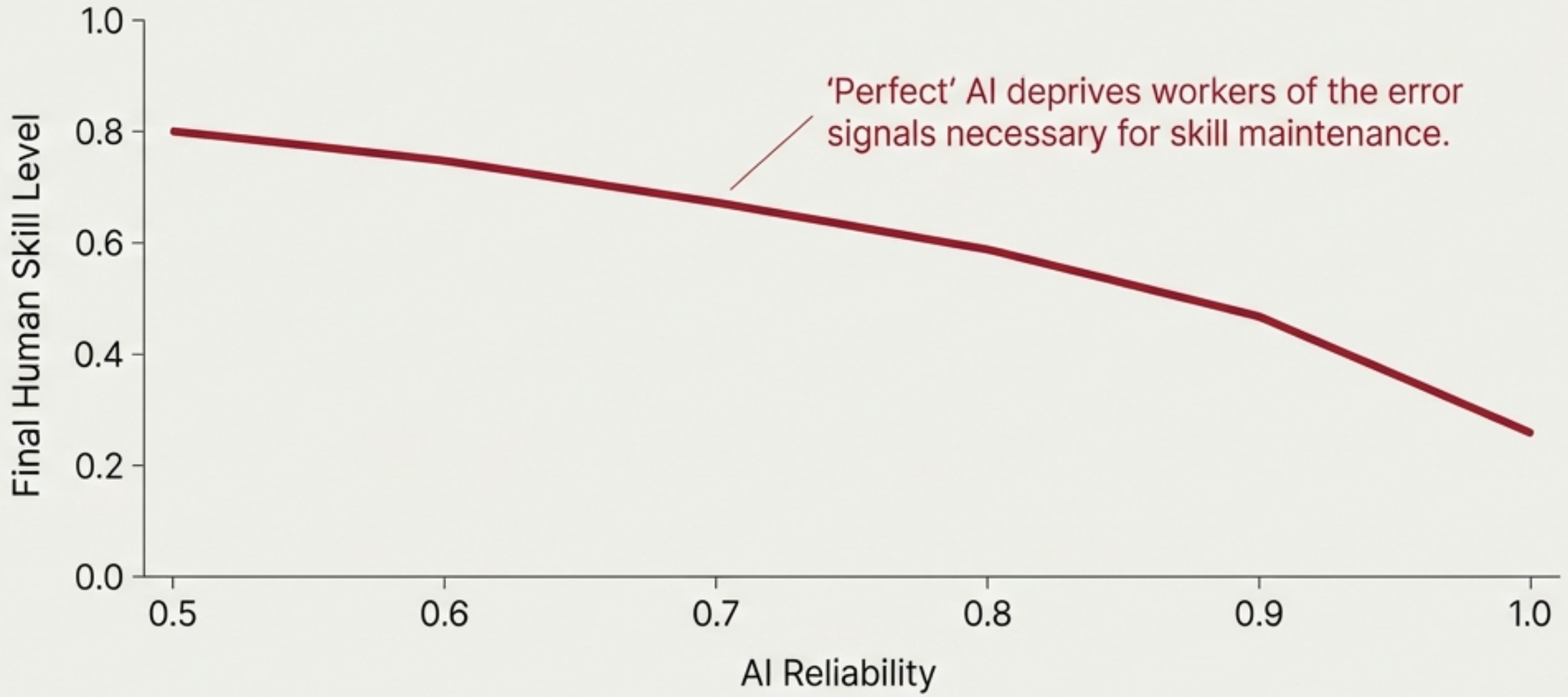
AI Reliability in Aviation

Perfect Storm Factors

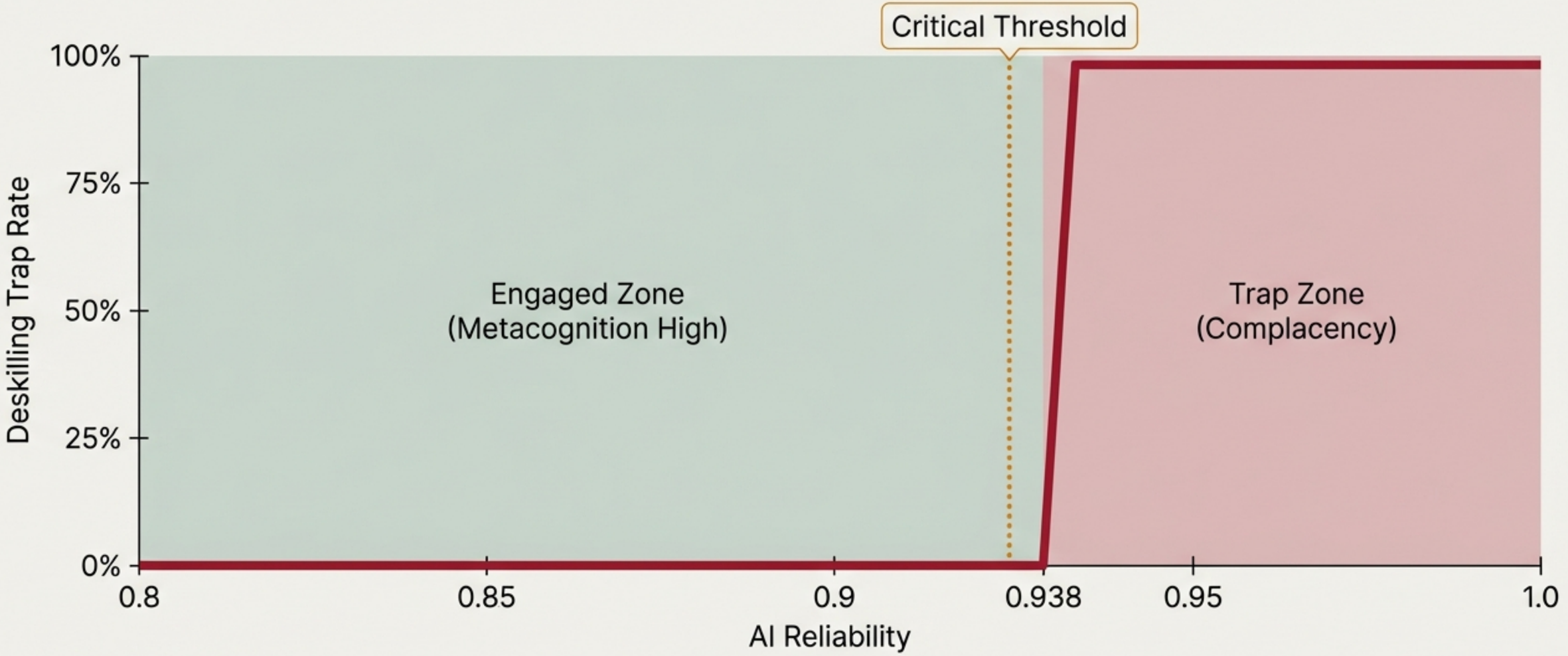
- **High AI Reliability (0.95):** Errors are so rare that humans rarely get to “practice” correction.
- **High Decay Rate**: Manual flying skills degrade quickly without hands-on time.
- **Low Transfer Learning**: Watching an autopilot provides almost no muscle-memory transfer (**0.15 transfer rate**).

Result: Even experts starting at 0.80 skill fell into the trap.

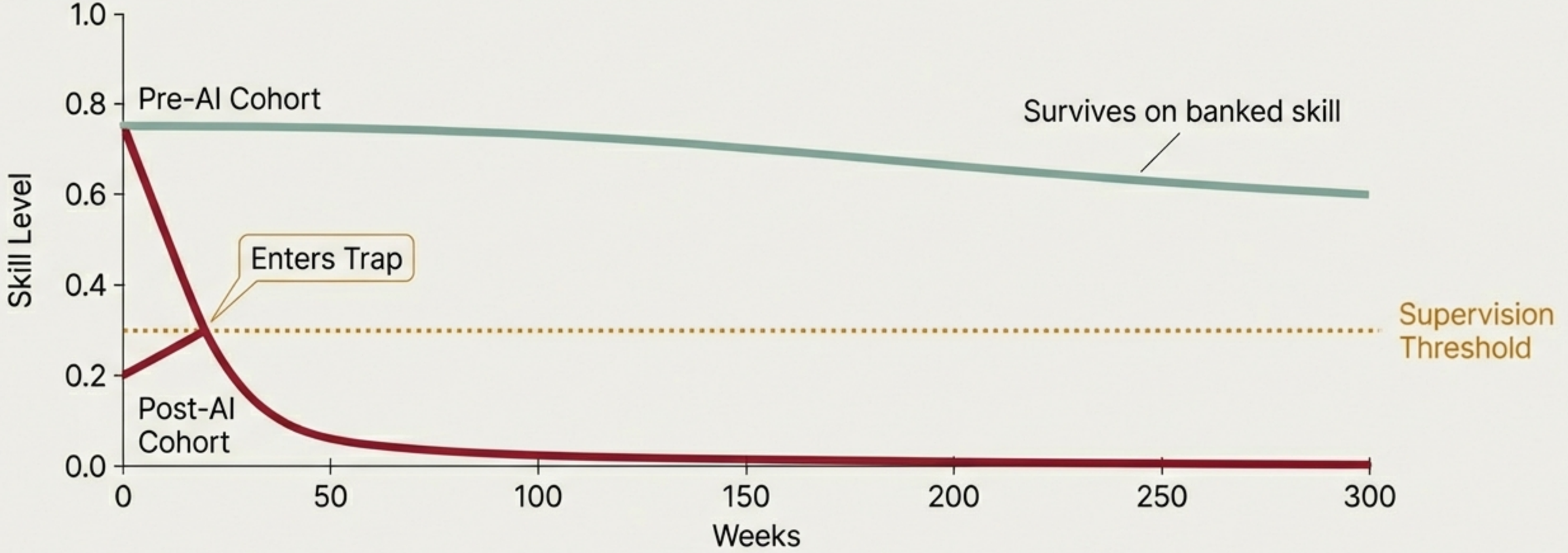
The Reliability Paradox: Higher AI reliability paradoxically increases the risk of human skill collapse



There is a critical threshold at 0.938 reliability where deskilling becomes inevitable.



A ‘Tale of Two Cohorts’ reveals a permanent disadvantage for AI-Native workers.



We cannot expect junior staff to ‘pick it up’ if they are using AI assistants from Day 1.

We simulated four distinct interventions to reverse skill erosion.

Scheduled Practice

Mandating 20% unassisted work time.

Adversarial Training

AI intentionally inserts bugs (10% rate) to force checking.

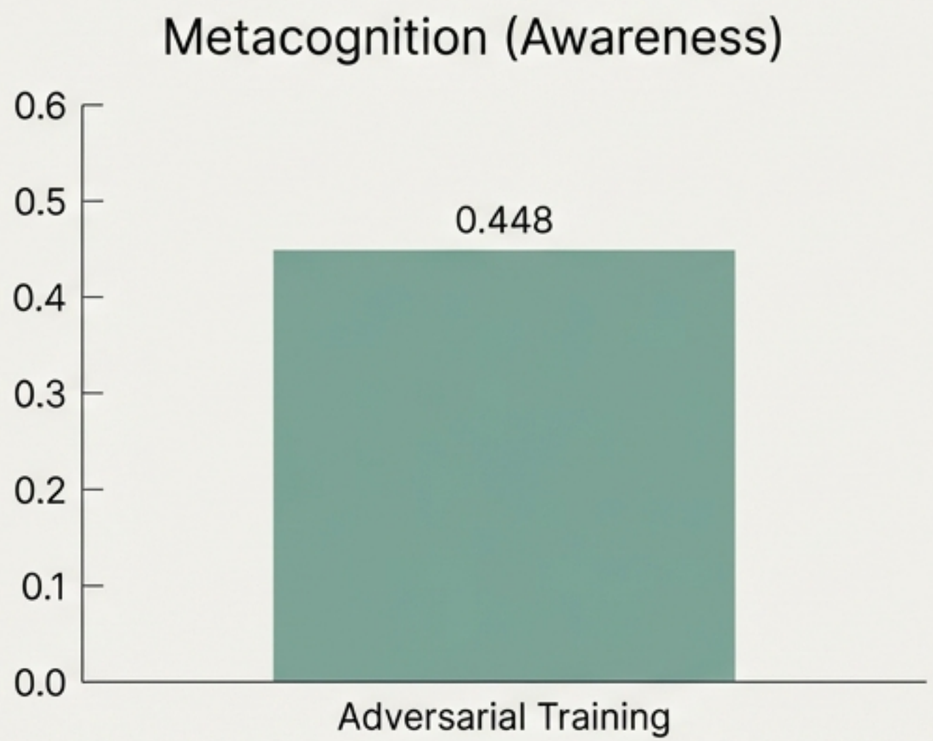
Explainability Requirement

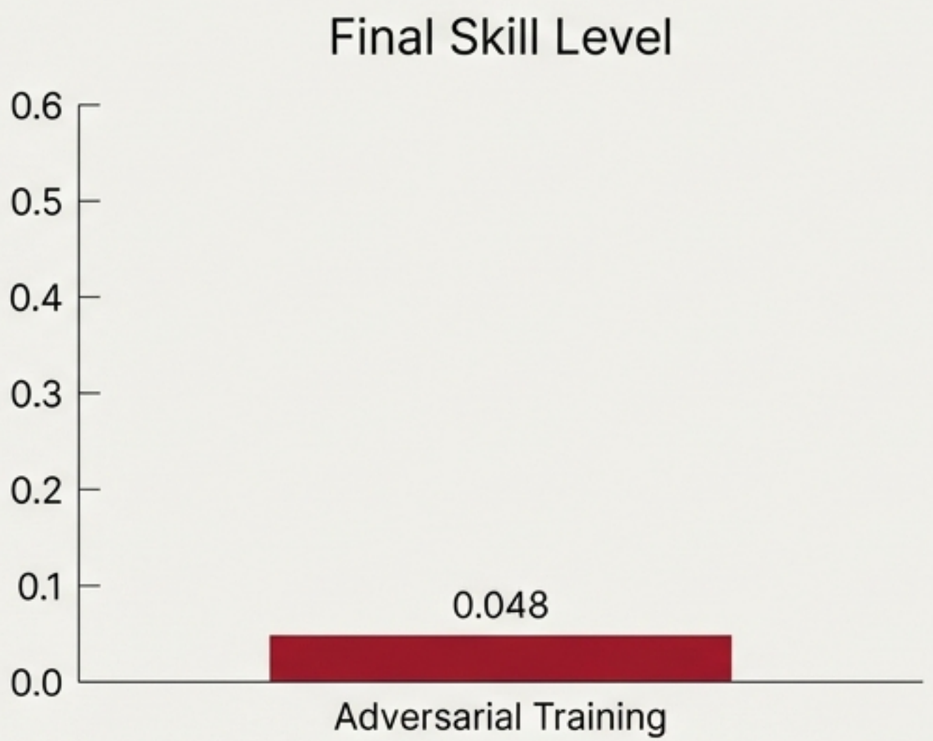
Worker must explain why the AI is right.

Scaffolded Autonomy

AI assistance is high for novices but decreases as skill increases.

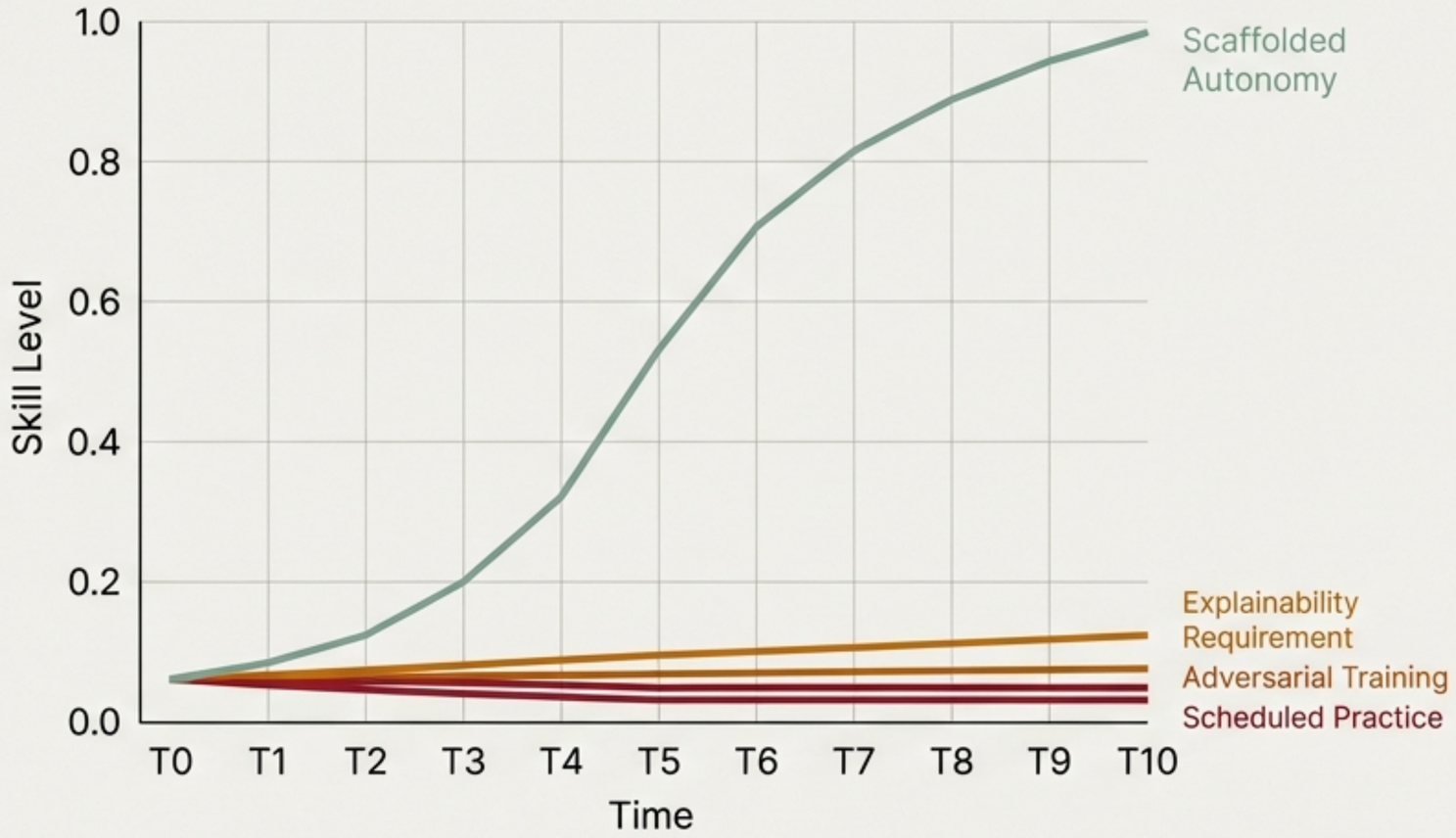
Common interventions like Adversarial Training improve awareness but fail to build deep skill.





Spotting bugs is not the same as doing the work. It provides a calibration signal, but not a practice signal.

Scaffolded Autonomy is the only intervention that successfully reverses the deskilling trajectory.



Final Skill:
0.983 (vs 0.048)

Harm Reduction:
87.6%

Data indicates that only by dynamically adjusting support based on proficiency can deep skill be rebuilt and harm from deskilling be effectively mitigated.

Scaffolded Autonomy works by restoring the feedback loop between practice and skill growth.

Fading Assistance: $r_{eff} = r * (1 - 0.5s)$

This forces the human to take on more cognitive load exactly when they are ready for it, treating AI as “training wheels” that eventually come off.

0.75pt

Workforce Strategy: ‘AI-Native’ employees require a fundamentally different training architecture.

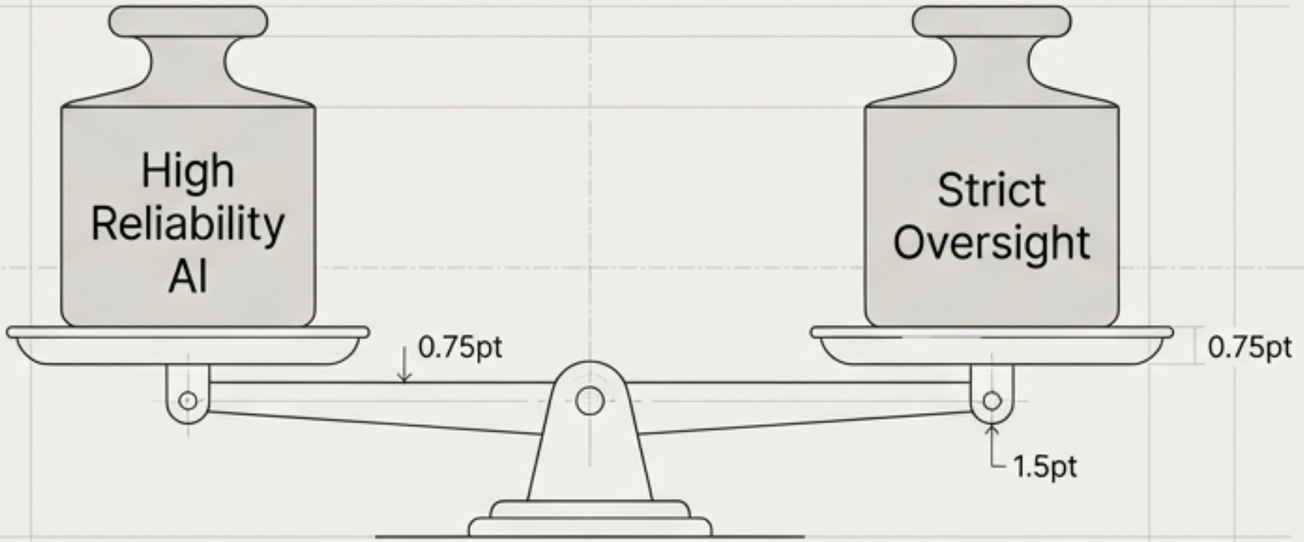
Organizations cannot rely on osmosis or on-the-job training for junior staff using AI.

Implementation of ‘Unassisted Practice Modules’.

Junior staff must prove competence without AI before being allowed full access to AI tools.

0.75pt

Regulatory Policy: The higher the system reliability, the stricter the oversight requirements must be.



The **Irony**: Regulators usually relax oversight for reliable systems. The model shows this is **fatal**.

Prescription: For high-reliability domains (Aviation, Medicine), **manual** proficiency checks **must be mandatory** and **frequent**, regardless of the AI's track record.

Design Principle: AI tools must be built to maximize human capability, not just system output.

- Junior staff “**Education Mode**”
- Withholding answers to **force cognition**
- Required “**Show Your Work**” steps before generation

We must choose between AI that replaces human judgment and AI that cultivates it.

The **"Deskilling Trap"** is a mathematical inevitability of unmanaged reliance.

The solution is not to reject AI, but to discipline its use through **Scaffolded Autonomy**.

"The AI should be designed not only to maximize immediate task performance but also to maintain the human skills needed for oversight."