A weaker model with payment tools is more autonomous than a frontier model with none.

Dr Peter McCann Strain
CTO, DPhil/PhD in AI from Oxford University

Swipe ››

—— THE OPERATING RITUAL

Teams grant autonomy to a model when they should grant it to an action class. A team that cannot say which rung it is running does not have an autonomy design; it has hope with credentials.

— THE REFRAME

Action classes (the kind of work and its consequence), not models.

THE OLD QUESTION

Is this model good enough?

↓ ↓ ↓

THE QUESTION THAT HOLDS UP

Which action may this system take, at what rung, under what controls?

— WORKED EXAMPLE, NOT A BENCHMARK

8% error rate*

Worked example, not a benchmark: in illustrative arithmetic, a 92 percent accurate system has an 8 percent error rate before detection, containment or severity is considered. At rung 1 that may be a drafting aid; at rung 4 it becomes delegated institutional action. The model did not change. The autonomy grant did.

SOURCE

Illustrative worked example in the essay; peer-reviewed support from Bainbridge's ironies-of-automation literature and METR's time-horizon work; the MIT AI Agent Index (2026 arXiv) is preprint context only.

CADENCE

The right rung is the highest rung at which four axes pass together, paired into three checks.

01 Check **reversibility** and **stakes**: how fast can a wrong action be undone?

02 Check **observability**: will the failure be noticed in time to contain it?

03 Check **accountability**: whose name is attached, and can that person act?

— THE ARTIFACT

Six rungs of autonomy, 0 to 5; the load-bearing boundary is between rung 3 and rung 4.

1. Chat
answers only

2. Drafted action
human sends

3. Approved action
human approves

4. Bounded autonomy
enumerated action space

5. Goal-directed
composes a path

6. Open-ended
highest scrutiny

Rung 0 Chat, 1 Drafted action, 2 Approved action, 3 Bounded autonomy, 4 Goal-directed autonomy, 5 Open-ended autonomy. The boundary that matters is between 3 and 4: enumerated action space below, runtime-composed above.

— PUT IT ON THE CALENDAR

Pick the last action your AI system took without a person at the keyboard. Write three sentences: action class, rung it actually operated at, recovery path. If any sentence is vague, lower the rung this week.

RITUAL DRIFT

Granting autonomy to the model and assuming each action inherits the right rung. Action classes climb the ladder independently. Score reversibility, observability and accountability per action, not per model, before approving the grant.

RUN THE OPERATING LOOP

The Autonomy Ladder

Read the full essay — the argument, the sources, the figures and a reader-ready working artifact.
Substack petermccannstrain.substack.com · Medium @peter.mccann.strain · LinkedIn peter-strain-dphil-15a607128
New essays twice weekly, 2 June – 21 July 2026.

Next: E22 — Stop Delegating. Start Architecting.

— THE STACK SO FAR

E21 · Essay 21 of 22 complete · Arc V: Operating model

YOU JUST ADDED

The Autonomy Ladder

STACK LAYER LIT UP

Permissions / Runtime evidence / Named owner

YOU CAN NOW ASK

assign autonomy by action class and rung.

NEXT

E22 asks who owns the ladder, the evidence, the exceptions and the authority to demote the system.

— THE ARTIFACT, CONTINUED

Six rungs of autonomy, 0 to 5; the load-bearing boundary is between rung 3 and rung 4.

THE REMAINING NODES

Rung 5 (Open-ended autonomy): the system sets or revises its own goals over time, with the human mostly outside the loop until escalation. It needs the heaviest containment.

Dr Peter McCann Strain

CTO, DPhil/PhD in AI from Oxford University

I build production AI systems and write about making agentic AI useful, inspectable, governable and safe enough for real work.

Follow on Substack for the full 22-essay series
petermccannstrain.substack.com