RESEARCH MANAGEMENT · AUTONOMOUS SCIENCE

Management of Autonomous Scientific Discovery

Opportunities and implications

Oleksiy Dzenis

Simon Kuznets Kharkiv National University of Economics (Ukraine)



Bohdan Yermakov

Imperial College London (United Kingdom)

IMPERIAL

IMPERIAL

THE PROBLEM

Generation is outpacing verification

2.2 million

new stable crystal structures predicted by one graph-network model

…of which **736** had been realised experimentally

3.02×

more papers published by scientists who adopt AI

…yet a **4.63%** contraction in the topics they study

Reviews cover algorithms and hardware. Missing: who, inside a research unit, decides *what* to delegate, *how far*, on *what evidence* and under *what controls*.

Sources: Merchant et al. (2023); Hao et al. (2026), 41.3 million papers · 02

IMPERIAL

RQ1 · THE OBJECT OF MANAGEMENT

Six delegable functions form the discovery loop

F1

Objective setting

Human: portfolio owner sets priorities and exclusions

F2

Hypothesis generation

Human: reviews plausibility and novelty claims

F3

Experiment design & prioritisation

Human: approves budgets and risk limits

F4

Execution

Human: maintains hardware, safety and supply

F5

Analysis & interpretation

Human: audits models and anomalies

F6 · added

Verification & communication

Human: signs off claims and owns accountability

Criteria: separable output · distinct cost of error · documented locus of delegation · 03

IMPERIAL

RQ1 · LEVELS OF AUTONOMY

Autonomy is decision authority, set per function

A0	**Manual**	No rights; the system is a tool
A1	**Assisted**	Proposes, ranks or recommends
A2	**Supervised**	Acts subject to consent at checkpoints
A3	**Delegated**	Acts without prior consent, within set limits
A4	**Self-directed**	Sets or revises the goals and limits

KEPT DISTINCT

D1 Authority - defines the level
D2 Capability - the ceiling
D3 Reliability - the warrant
D4 Oversight - the complement

$$\Delta_k = a_k - w_k$$

Delegation gap: granted minus warranted authority. Δ > 0 = over-delegation.

Built from four transferable decision rights: propose · act with consent · act within limits · set the limits · 04

RQ3 · FORMALISATION

The loop runs at the speed of its scarcest capacity

TESTING RATE · EQ. (3)

$$\Lambda = min(generate, execute, analyse)$$

Tests per period are bounded by the weakest stage, not the strongest.

VERIFIED DISCOVERY RATE · EQ. (4)

$$VDR_t = v \cdot min(h \cdot \Lambda + VD_{t-1}, verify)$$

Only verified claims count, capped by verification capacity; v is the share that survives.

NEW · VERIFICATION DEBT · EQ. (5)

$$VD_t = max(0, VD_{t-1} + h \cdot \Lambda - verify)$$

Unverified claims carry over as a liability for those who build on them.

NEW · DIVERSITY-ADJUSTED RATE · EQ. (6)

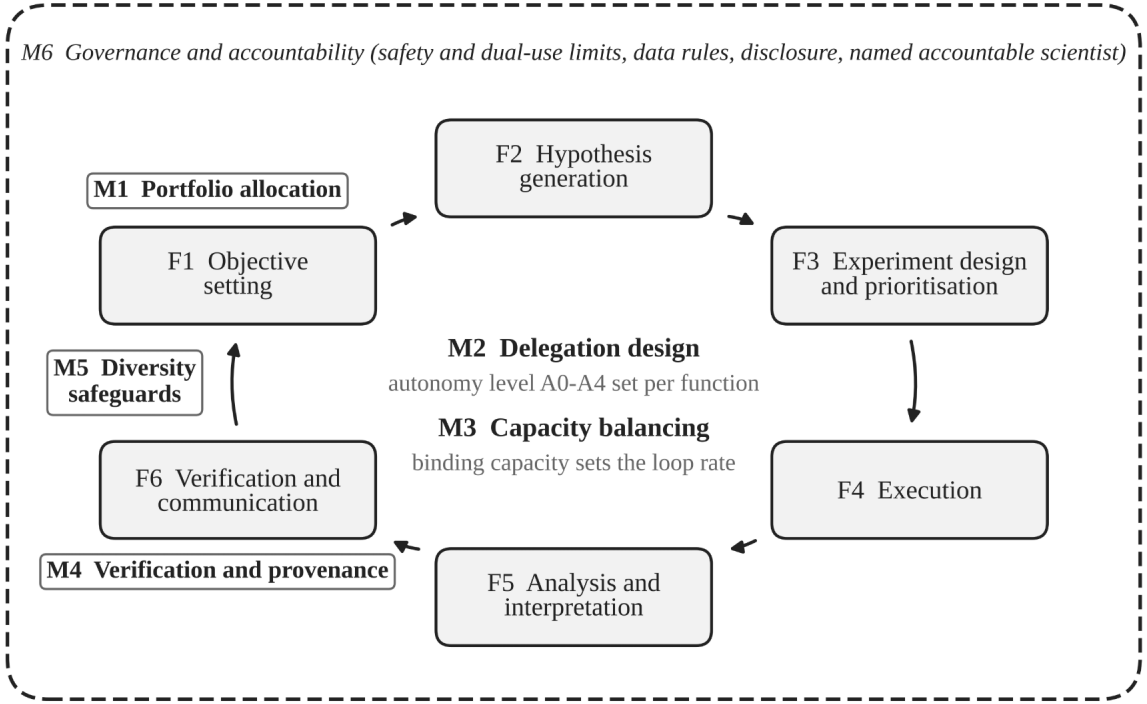
$$Q = VDR \cdot (1 - HHI)$$

Concentration across topics (Herfindahl index) discounts output.

verify = claims that can be independently verified per period · investment is appraised by cost per verified claim, Eq. (7) · 05

RQ2 · OWNER, TRIGGER, INDICATOR, FAILURE MODE

Six mechanisms manage the loop



MECHANISM → HOW IT FAILS

M1 Portfolio allocation
Objective drift to what is easy

M2 Delegation design
Implicit delegation

M3 Capacity balancing
Funding generation while verification binds

M4 Verification & provenance
Claims released before verification

M5 Diversity safeguards
Efficient monoculture

M6 Governance & accountability
Governance applied after the fact

Fig. 2 of the paper · each with an owner, trigger, indicator and failure mode (Table 6) · 06

IMPERIAL

RQ2 · MECHANISMS ACROSS AUTONOMY LEVELS

Governance changes form as autonomy rises

A0-A2

Control precedes action

M2 records who approves what and sets consent checkpoints

M4 checks results people have already accepted

M6 works through sign-off and records each consent

A3

Prior consent disappears

M2 sets limits, risk classes, stop conditions and provenance records

M4 re-analyses a sample of claims before release, scaled by cost of error and w_k

M6 escalates out-of-limit actions to a named, accountable person

A4

Goals become a mandate

M1 grants objectives as a mandate on a fixed review cycle (F1, F3); M2 for other functions

M5 monitors drift of self-set objectives

M6 institution grants, renews or withdraws the mandate

> M3 matters more at A3-A4, where generation can outgrow verification. Raise a function's level only once the mechanisms of that level are in place.

HYPOTHETICAL ILLUSTRATION · THREE SCENARIOS

Most raw claims, least verified knowledge

Hypothetical scenario	Testing bound by	Raw claims / wk	Verified (VDR)	Debt / wk	Q
A · University group, self-driving lab	Generation	4	2.4	+1	1.92
B · Corporate unit, robotic screening	Execution	6	2.8	+2	1.82
C · AI-native start-up, generative agents	Execution	10 (most)	1.0 (least)	+8	0.40

> Ranked by raw claims, C comes first; by verified output, last. At equal operating and rework cost, C has the lowest cost per raw claim but the **highest cost per verified claim**.

Weekly values are assumed for demonstration and are not empirical observations (Table 8) · 08

FALSIFIABLE BY DESIGN

Five testable propositions

P1	Once execution or analysis binds, more generative autonomy does **not** raise the verified discovery rate.
P2	Verification debt at release predicts later corrections and failed replications. **Key test.**
P3	More autonomy in F1–F3 without diversity safeguards brings rising portfolio concentration.
P4	Explicit delegation design means shorter cycle times and fewer late governance escalations.
P5	Functions with a positive delegation gap contribute disproportionately to debt and corrections.

IMPERIAL

THE TAKEAWAY

The lever is rarely generative capacity. It is the capacity to verify, diversify and account for what autonomous systems produce.

1

Profile the loop

Authority, reliability, capacities and gaps per function

2

Relieve the binding capacity

Rank investments by the rise in VDR per unit of cost

3

Make delegation explicit and calibrated

Raise a level only once its mechanisms are in place

4

Encode governance and diversity as preconditions

Escalate only novel cases and level changes

$$CVC = (K + R) / VDR$$

Cost per verified claim, Eq. (7). Operating cost K plus rework cost R, per verified claim. An investment improves efficiency if it lowers CVC, not if it raises throughput.

Thank you · oleksiy.dzenis@hneu.net · bohdan.yermakov25@imperial.ac.uk