具身化 AI 自主研究閉環

v0.1 / 2026-07-30 / Neo.K / 系統統合論文／具身自主研究閉環規格

證據狀態：E0——形式模型、治理架構、驗證協議與 MVP 路線

https://thisoneisneok.com/html/papers/stdi-10.html

SECTION

從假說生成、物理實驗到證據判定與概念修正

Embodied AI Autonomous Research Loop: From Hypothesis Generation and Physical Experimentation to Evidence Adjudication and Concept Revision

系列：《時空域支配智能》系列第 10 篇

文件編號： EML-STDI-EARC-2026-v0.1

架構名稱： EARC｜Embodied Autonomous Research Cycle

作者： Neo.K

協作整理： Aletheia（阿萊）

機構： EveMissLab／一言諾科技有限公司

日期： 2026 年 7 月 30 日

文件類型： 系統統合論文／具身自主研究閉環規格

證據成熟度： E0——形式模型、治理架構、驗證協議與 MVP 路線

公開狀態： 私人研究稿；公開前應經 EML-CF 的 IP Gate、來源、安全、倫理與證據成熟度審查

上位理論： STDI｜時空域支配智能

直接前序：

- 《時空間支配型 AI：從單體具身智能到持續性時空域治理》
- 《超靈的物理化：從 O-Chip 維度代理人到分布式具身主體》
- 《Oversoul Station Fabric：固定站、移動站與虛擬站的分布式具身網路》
- 《持續性指揮控制區：AI 如何佔據、維持並安全解除一個物理時空域》
- 《語義即物理路由：從資料流治理到物料、能源、站點與行動流治理》
- 《具身即佔域，對齊即能力：分布式身體的時序容量、協調稅與規模邊界》
- 《中央主權、地方自治與動態不動點中央》
- 《連線不是纜線：有線、無線、光學與離線任務包的混合站網》
- 《站點化世界模型：物體、區域、事件、權限與可能行動的共同物理世界表示》

SECTION

摘要

前九篇已建立一個持續性超靈如何取得多個身體、組織站點網、維持物理時空域、編譯物料與能源流、計算對齊容量、分配中央與地方權限、跨越異質鏈路，並維護可區分觀測、推論、模擬和提交狀態的共同世界模型。本文將這些元件第一次閉合成完整研究系統，提出 EARC｜Embodied Autonomous Research Cycle（具身化 AI 自主研究循環）。

EARC 的目的不是把人類科學家從每一個研究階段完全刪除，也不是讓大型語言模型以自然語言直接操作實驗室。它要處理的是一個更嚴格的工程問題：

> 一個 AI 研究系統如何從概念產品或科學問題出發，建立可反駁的假說，編譯可執行實驗，調度真實物理站點，取得帶來源的量測證據，判定證據支持、削弱或無法區分哪些主張，再把結果回饋至下一輪假說、設計、產品與智財路由？

本文定義研究狀態：

$$
\begin{aligned}
&\text{cal-R}\boxtimes \\
&= \\
&(\\
&K\boxtimes, \\
&H\boxtimes,
\end{aligned}
$$

其中：

- K⊠：文獻、先驗知識與既有證據；
- H⊠：候選假說與競爭解釋；
- X⊠：可執行實驗與干預集合；
- W⊠：站點化世界模型；
- D⊠：新觀測、量測與實驗資料；
- E⊠：證據圖、校準、來源與反證；
- M⊠：物理、統計、因果與生成模型；
- B⊠：時間、材料、能源、風險和算力預算；
- G⊠：安全、倫理、IP、公開與人類治理規則。

  EARC 的閉環為：

> Concept／Question
> →
> Evidence-Grounded Scope
> →
> Hypothesis Contract
> ;
> →
> Experiment Portfolio
> →
> DomainIR Compilation
> →
> Simulation／Safety Gate
> ;
> →
> OSF Physical Execution
> →
> Measurement and Provenance
> ;

本文特別區分四種經常被混為一談的自動化：

- 程序自動化：人類已決定完整程序，機器負責重複執行；
- 實驗最佳化閉環：AI 在既定參數空間中選擇下一個實驗；
- 假說判別閉環：AI 選擇最能區分競爭解釋的實驗；
- 研究主線閉環：系統可以修正問題、假說、模型、實驗表示、證據要求和停止條件。

  現有自主實驗室已證明閉環物理研究的多個局部能力。移動機器人化學家曾在八天內執行 688 次實驗，於十變量空間中以批次貝葉斯搜尋找到活性高於初始配方六倍的光催化混合物；A-Lab 將機器人、計算材料資料、文字挖掘合成啟發式、量測解讀和主動學習整合，在有限無機材料目標集合中以最少人力進行自主合成；Coscientist 展示多 LLM 系統可搜尋文獻與設備文件、規劃化學程序並透過 API 操作雲端或實體實驗設備。這些成果證明 AI 可以分析、選擇和執行下一輪實驗。

  但 2026 年的自主實驗室管理研究仍指出，成功系統多半是針對狹窄研究活動、少量工具與明確目標打造的客製閉環。要從「自駕實驗」進入「自主管理複雜研究主線」，還需要跨工具、跨站點、跨時間、跨代理、跨證據成熟度與跨組織邊界的治理系統。本文將前九篇視為這個缺口的基礎設施。

  EARC 不允許研究閉環退化成「AI 生成假說、AI 選擇支持自己的實驗、AI 解讀結果、AI 宣布成功」的高速自我確認。因此本文建立：

- 競爭假說；
- 預先指定的反駁條件；
- 正負對照；
- 盲測和保留集；
- 正交量測；
- 反對者／驗證 Agent；
- 證據品質權重；
- 負結果與失敗保存；
- 獨立重複；
- 人類風險與主張提交權；

- 明確停止和棄權條件。

  實驗選擇不只追求性能最大化，而可根據資訊增益、假說區分度、重現價值、預期產品價值、成本與風險選擇：

> x^*
> =
> argmax_(x∈cal-X_(exec))
> [
> α I(H;Ymid x)
> +
> β V_(discriminate)(x)
> +
> γ V_(replicate)(x)
> +
> η V_(product)(x)
> -
> λ C(x)
> -
> μ R(x)
> -
> ν T(x)
> -
> ξ X_(IP)(x)
>].

本文提出六級證據成熟度：

E0 概念與形式命題
E1 模擬或計算支持
E2 單次受控物理觀測
E3 校準、重複與反證後的內部證據
E4 跨站、跨批次或獨立重現
E5 長期、外部或實際運作環境驗證

一個概念產品可以在 E0 階段被保存和發展，但不能以 E0 語言宣稱 E4 或 E5 的現實可靠性。EML-CF 依證據與智財成熟度將結果分流至私人研究、專利優先、營業秘密、內部工程、合作驗證、防禦公開、開源或停止。

本文的核心命題為：

> 自主研究不是讓 AI 代替人類按下所有按鈕，而是建立一個能把想法反覆暴露於物理反饋、反例、失敗、校準和責任審查中的可停止閉環。

關鍵詞： 具身化 AI、自主研究閉環、EARC、自主實驗室、假說契約、實驗設計、主動學習、物理驗證、證據判定、反證、重現、負結果、EML-CF、DomainIR、OSF、世界模型、人類監督、智財路由

前九篇分別解決了「身體、域、站網、時間、路由、容量、治理、連線和世界」。第十篇要回答的是：

> 這些基礎設施如何真正縮短概念產品和科學命題從提出到被物理世界淘汰、修正或支持的時間？

EARC 不是另一個獨立實驗室平台，而是：

> 【EML-CF
> +
> SWM
> +
> SPR／DomainIR
> +
> OSF
> +
> PCD
> +
> DFC
> +
> HLF
> +
> Evidence Engine】

的研究控制閉環。

其直接目標是把：

> 概念生成
> →
> 等待研究室
> →

改寫成：

> 概念
> →
> 可反駁主張
> →
> 可編譯實驗
> →
> 可治理物理執行
> →
> 可追溯證據.

SECTION

1.1 程序自動化

人類事先決定：

- 做什麼；
- 順序；
- 參數；
- 停止；
- 結果解讀。

設備只重複執行。

這能提高：

- 速度；
- 一致性；
- 吞吐；
- 可重現性。

但不是自主研究。

SECTION

1.2 參數最佳化閉環

系統在固定空間中選擇：

> $$x_{t+1} = \pi(D_{\leq t}).$$

常見目標：

- 最大產率；
- 最低成本；
- 最佳材料性質；
- 最佳反應條件。

這是現有 self-driving laboratory 最成熟的形式之一。

SECTION

1.3 假說判別閉環

研究目的不是找最大值，而是區分：

> $$H_1, H_2, \ldots, H_n.$$

下一個實驗應最大化：

- 預測差異；
- 因果辨識；
- 模型淘汰；

- 可解釋性。

SECTION

1.4 研究主線閉環

系統允許修改：

- 問題邊界；
- 假說集合；
- 實驗表示；
- 儀器；

  -證據要求；
- 研究停止條件；
- 概念產品規格。

  這是 EARC 的目標，但也具有最高治理風險。

SECTION

2.1 移動機器人可以使用傳統實驗室

移動機器人化學家展示：

- 自主移動；
- 操作多個既有儀器；
- 八天持續工作；
- 688 次實驗；
- 十變量搜尋；
- 批次貝葉斯最佳化。

其重要意義是：

> 不一定要把整個實驗室改造成一台封閉機器；也可以讓具身代理在傳統設備間移動。

SECTION

2.2 理論、文獻、機器人與主動學習可以閉合

A-Lab 整合：

- 計算材料資料；
- 文字挖掘程序；
- 機器人合成；
- XRD 等量測解讀；
- 失敗後的主動學習。

這證明閉環可以不只做參數掃描，也能把計算預測和實驗失敗回饋至下一輪程序。

SECTION

2.3 LLM 可以成為實驗規劃和工具調用層

Coscientist 展示：

- 網路與文獻搜尋；
- 設備文件理解；
- 程序規劃；
- Python 計算；
- 實驗 API；
- 多模組協調。

但這類架構也顯示：

> LLM 的高階語言推理必須與文件、設備 API、低階控制和量測工具分離。

SECTION

2.4 分布式實驗可以非同步閉環

跨實驗室、跨設備和跨地點的自動發現已開始出現。這支持 OSF、HLF 和 PCD 的路線：研究閉環不必綁在同一房間，也不必要求所有站點同時在線。

SECTION

2.5 前沿系統開始宣稱研究全週期

2026 年已有預印本系統宣稱在量子材料或光學平台中，從假說、規劃、實驗執行、分析到新機制驗證形成較完整閉環。這些是重要的前沿訊號，但仍應依：

- 是否經同行評議；
- 是否可重現；
- 是否存在人類隱性介入；
- 新穎性判定；
- 統計與量測證據；
- 原始資料和審計；

分級看待，而不能直接作為「通用自主科學家已完成」的證明。

SECTION

3.1 多數系統是研究活動，不是完整實驗室治理

成功案例常具有：

- 狹窄問題；

- 固定參數空間；
- 少量工具；
- 客製整合；
- 已知量測；
- 明確目標函數。

跨多個研究主線、工具、材料、模型、IP 和人類參與者的實驗室管理，仍是更高層問題。

SECTION

3.2 閉環容易把最佳化當科學

找到：

> x^*=argmax f(x)

不等於理解：

- 為什麼；
- 機制是什麼；
- 是否可外推；
- 是否可重現；
- 是否存在替代解釋。

SECTION

3.3 物理完成不等於證據成立

機械臂完成動作後，仍可能：

- 樣本身份錯；

- 儀器未校準；
- 對照缺失；
- 數據處理錯；
- 世界紀元過期；
- AI 選擇性忽略負結果。

SECTION

3.4 自我改進可能放大錯誤

若同一系統：

- 生成假說；
- 選擇實驗；
- 決定品質；
- 解讀結果；
- 決定是否成功；

就可能形成閉環偏差。

SECTION

3.5 研究價值不等於儀器利用率

高吞吐可能產生大量低資訊實驗。EARC 必須最佳化：

- 資訊；
- 區分；
- 反例；
- 可重現；

- 產品價值；

  而不只是設備忙碌時間。

  定義：

> cal-R□
> =
> (
> K□,
> H□,
> X□,
> W□,
> D□,
> E□,
> M□,
> B□,
> G□
>).

SECTION

4.1 K□：知識狀態

包括：

- 文獻；
- 專利；
- 內部筆記；
- 既有數據；
- 失敗；
- 方法；

- 先驗。

所有內容攜帶來源與公開狀態。

SECTION

4.2 H□：假說狀態

包括：

- 主假說；
- 零假說；
- 競爭解釋；
- 工程假設；
- 可反駁預測；
- 未知。

SECTION

4.3 X□：實驗候選

包括：

- 模擬；
- 計算；
- 物理；
- 觀測；
- 破壞性；
- 非破壞性；
- 重現；

- 反證。

SECTION

4.4 W□：物理世界

由 SWM 提供。

SECTION

4.5 D□：資料

原始量測、衍生結果、圖像、事件和日誌。

SECTION

4.6 E□：證據

不是所有資料都自動成為證據。

SECTION

4.7 M□：模型

- 理論；
- 統計；
- 因果；
- 模擬；
- 生成式世界模型；
- 儀器模型。

SECTION

4.8 B□：預算

時間、材料、能源、風險、金錢、算力和人力。

SECTION

4.9 G□：治理

安全、倫理、IP、合作、公開和停止權。

SECTION

5.1 假說不能只是一句靈感

定義：

> h
> =
> (
> C,
> S,
> V,
> P,
> F,
> A,
> E,
> R
>),

其中：

- C：核心主張；
- S：適用範圍；

- V：可觀測變量；
- P：預測；
- F：反駁條件；
- A：替代解釋；
- E：最低證據；
- R：風險、IP 與倫理限制。

SECTION

5.2 工程概念也能建立假說

例如：

> 幾何表面結構 A 在相同材料、體積和環境下，比基準結構 B 提供更低的穩態熱阻。

必須補出：

- 相同條件；
- 測量定義；
- 誤差容限；
- 樣本差異；
- 反例；
- 最低改善幅度。

SECTION

5.3 競爭假說

至少包含：

- H⊠：無顯著改善；

- H□：幾何機制造成改善；
- H□：改善來自製造誤差或材料差；
- H□：量測系統偏差。

SECTION

5.4 不允許事後移動門檻

證據門檻、主要指標和停止條件應在執行前保存版本。

研究不是單一假說，而是主張圖：

> G_H
> =
> (
> V_H,
> E_(support),
> E_(contradict),
> E_(depend),
> E_(refine)
>).

SECTION

6.1 主張類型

- 存在；
- 比較；
- 機制；
- 因果；
- 預測；

- 工程可行；
- 安全；
- 經濟；
- 可擴展。

SECTION

6.2 主張依賴

「產品可行」可能依賴：

- 物理效果存在；
- 效果可重複；
- 製造可行；
- 成本可接受；
- 安全可管理；
- 無關鍵專利阻礙。

SECTION

6.3 結論不能越過依賴鏈

單次物理效果成立，不能直接推出商業化成功。

SECTION

7.1 Experiment Portfolio

EARC 為每輪建立：

```
cal-X⊠
=
{
```

SECTION

7.2 探索

尋找：

- 未知現象；
- 參數區域；
- 異常；
- 候選機制。

SECTION

7.3 利用

提高性能或縮小最佳區域。

SECTION

7.4 反證

主動尋找最可能推翻主張的條件。

SECTION

7.5 重現

重複關鍵結果。

SECTION

7.6 校準

量測系統和模型可信度。

SECTION

7.7 壓力測試

測試：

- 邊界；
- 故障；
- 長時間；
- 批次；
- 環境變化。

SECTION

8.1 純最佳化

> x^*
> =
> argmax_x
> bb-E[f(x)].

只適合明確最佳化問題。

SECTION

8.2 不確定性降低

> x^*
> =
> argmax_x
> I(H;Ymid x).

SECTION

8.3 假說判別

選擇使不同假說預測差距最大的實驗：

> x^*
> =
> argmax_x
> Divergence
> (
> p(ymid H⊠,x),
> p(ymid H⊠,x)
>).

SECTION

8.4 綜合效用

> U(x)
> =
> α I(H;Ymid x)
> +
> β V_(discriminate)(x)
> +
> γ V_(replicate)(x)
> ;
> +
> η V_(product)(x)
> +
> θ V_(anomaly)(x)
> -
> λ C(x)
> ;
> -

SECTION

8.5 可執行集合

只在：

> cal-X_(exec)
> =
> {
> x
> mid
> Compiled
> ∧
> Safe
> ∧
> Authorized
> ∧
> EvidenceReady
> }

中選擇。

每個候選實驗先經 SPR：

> Experiment Design
> →
> DomainIR
> →
> OSF Binding.

檢查：

- 樣本；
- 工具；

- 儀器；
- 路徑；
- 能源；
- 時間；
- 權限；
- 世界紀元；
- 證據；
- 補償。

SECTION

9.1 不可編譯實驗

可能是：

- 不存在能力；
- 樣本不足；
- 不可量測；
- 幾何不可達；
- 安全不允許；
- 缺少對照；
- 成本超限；
- IP 不可外送。

不可編譯不表示假說錯誤，只表示當前基礎設施不能測試。

SECTION

10.1 模擬目的

不是替代實驗，而是：

- 檢查程序；
- 預測風險；
- 篩除明顯無效條件；
- 估計量測範圍；
- 產生失敗模式；
- 驗證機械可達性。

SECTION

10.2 多模型交叉

使用：

- 規則；
- 數值模擬；
- 數位孿生；
- 統計模型；
- 世界基礎模型；
- 因果模型。

SECTION

10.3 模型分歧也是資訊

若模型對同一實驗預測高度分歧，該實驗可能具有高辨識價值。

SECTION

10.4 模擬通過不是物理批准

> Simulation Pass
> not⇒
> Physical Authorization.

SECTION

11.1 風險分級

R0 純計算或只讀觀測
R1 低能量、可逆、無危險材料
R2 有設備、樣本或人員風險
R3 高能量、危險材料、不可逆
R4 法律、人體、環境或公共重大風險

SECTION

11.2 批准

- R0：自動；
- R1：預先政策內自動；
- R2：中央與指定人類；
- R3：明確人類批准和現場安全；
- R4：專門治理與外部合規。

SECTION

11.3 IP Gate

在外部站點、雲端或公開系統參與前檢查：

- 是否專利候選；
- 是否營業秘密；
- 是否已公開；
- 是否有 NDA；
- 哪些數據可外送；
- 哪些模型可下載。

SECTION

11.4 人類保留的決定

- 問題是否值得；
- 風險是否接受；
- 不確定結果如何解讀；
- 是否聲稱新發現；
- 是否申請專利；
- 是否公開；
- 是否停止。

EARC 不是單一速度的迴圈，而是巢狀控制。

SECTION

L0：硬體閉環

微秒至毫秒：

- servo；

- 互鎖；
- 急停。

SECTION

L1：站點閉環

毫秒至分鐘：

- 局部視覺；
- 路徑；
- 工具；
- 重試。

SECTION

L2：實驗閉環

分鐘至天：

- 一次程序；
- 量測；
- 分支；
- 補償。

SECTION

L3：研究主線閉環

天至月：

- 假說；

- 實驗組合；
- 模型更新；
- 重現。

SECTION

L4：概念產品組合閉環

月以上：

- 產品規格；
- 專利；
- 原型；
- 停止；
- 公開；
- 商業化。

高層不能直接取代低層。

SECTION

13.1 Measurement Contract

$$
\begin{aligned}
&m \\
&= \\
&(\\
&\text{quantity}, \\
&\text{instrument}, \\
&\text{range}, \\
&\text{resolution}, \\
&\text{uncertainty}, \\
&\text{calibration}, \\
&\text{sample},
\end{aligned}
$$

SECTION

13.2 必須保存

- 原始資料；
- 儀器設定；
- 校準；
- 樣本身份；
- 環境；
- 時間；
- 軟體版本；
- 失敗；
- 人類介入。

SECTION

13.3 量測不是證據的全部

證據品質：

> q_e
> =
> q_(cal)
> q_(prov)
> q_(repro)
> q_(relevance)
> q_(independence)
> q_(integrity).

任一關鍵因子接近零，整體證據權重下降。

SECTION

14.1 三分判定不足

結果不只分為支持和反對。

SUPPORTS
WEAKENS
FALSIFIES
INCONCLUSIVE
CONFOUNDED
MEASUREMENT_FAILED
OUT_OF_SCOPE
REQUIRES_REPLICATION

SECTION

14.2 貝葉斯更新

在適用時：

> p(H⊠mid D)
> ∝
> p(Dmid H⊠)p(H⊠).

但 EARC 不要求所有科學都被壓成單一貝葉斯模型。

SECTION

14.3 頻率、效應與實質意義

同時報告：

- 效應量；
- 不確定性；
- 統計判定；

- 工程最低重要差；
- 批次差異；
- 模型敏感度。

SECTION

14.4 不可辨識

若不同假說在現有量測下產生近似結果：

> p(Dmid H⊠)
> ≈
> p(Dmid H⊠),

系統應輸出「不可區分」，而不是任意選一個。

SECTION

15.1 角色分離

至少區分：

- Hypothesis Generator；
- Experiment Designer；
- Safety／Feasibility Verifier；
- Physical Executor；
- Evidence Critic；
- Replication Planner；
- Human Committer。

SECTION

15.2 反對者 Agent

專門尋找：

- 替代解釋；
- 缺少對照；
- 數據洩漏；
- 指標遊戲；
- 儀器偏差；
- 選擇性報告；
- 失敗被忽略。

SECTION

15.3 預註冊

在實驗前保存：

- 主假說；
- 指標；
- 樣本數；
- 排除規則；
- 停止條件；
- 分析版本。

SECTION

15.4 盲測

可對：

- 樣本身份；
- 對照；
- 分析標籤；
- 模型版本；

  做適當盲化。

SECTION

15.5 保留集

部分樣本或條件不進入最佳化模型，只用於最後驗證。

SECTION

15.6 負結果不可刪除

失敗和零結果是：

- 模型更新；
- 風險邊界；
- 專利判斷；
- 重複避免；

  的重要資產。

  研究 Agent 的每個關鍵步驟可插入驗證探針：

- 引用是否支持主張；
- 數據是否存在；
- 實驗是否依預先版本；

- 分析是否更換指標；
- 結論是否超出證據成熟度；
- 是否漏報反例；
- 是否存在 IP 暴露。

驗證探針的目的，是把：

> AI 說它完成了

改成：

> 這裡是它做了什麼、依據什麼、哪些地方仍未被證明。

SECTION

17.1 可自動更新

- 參數後驗；
- 失敗概率；
- 站點效能；
- 模型權重；
- 實驗優先級；
- 候選假說可信度。

SECTION

17.2 不可自動改寫

- 安全憲法；
- 法律；

- 人類權利；
- 公開權；
- 原始事件；
- 既有負結果；
- 已提交證據。

SECTION

17.3 模型升級

升級前：

- 保存舊模型；
- 建立回歸測試；
- 重放歷史案例；
- 影子模式；
- 比較決策；
- 人類批准必要變更。

SECTION

18.1 Failure Record

> f
> =
> (
> stage,
> cause,
> physical_effect,
> recoverability,
> evidence,

SECTION

18.2 失敗分類

- 假說錯；
- 實驗不可辨識；
- 編譯錯；
- 能力錯配；
- 機械故障；
- 量測失效；
- 數據分析錯；
- 世界狀態錯；
- 治理阻擋；
- 資源不足。

SECTION

18.3 不把所有失敗餵給同一模型

不同失敗應更新不同元件：

- 物理模型；
- 能力證書；
- 路由；
- 世界模型；
- 實驗設計；
- 假說；

- 安全政策。

SECTION

E0：概念命題

- 形式化；
- 文獻；
- 邏輯一致；
- 尚未物理測試。

SECTION

E1：計算／模擬

- 數值；
- 數位孿生；
- 模型預測；
- 不等同實驗。

SECTION

E2：單次物理觀測

- 可證明現象曾出現；
- 尚不能證明穩定、可重現。

SECTION

E3：內部受控證據

- 校準；
- 對照；
- 重複；
- 反證；
- 完整來源。

SECTION

E4：跨站或獨立重現

- 不同設備；
- 不同批次；
- 不同 Agent；
- 或外部合作方。

SECTION

E5：實際環境驗證

- 長時間；
- 真實負載；
- 使用者；
- 製造；
- 現場。

SECTION

19.1 成熟度不單調自動上升

新反例可使：

> E4→ E2

或使原主張縮小適用範圍。

自主研究系統必須知道何時不再繼續。

SECTION

20.1 成功停止

- 主張達到證據門檻；
- 工程規格達標；
- 可進入下一產品階段。

SECTION

20.2 反駁停止

- 核心主張被反證；
- 不再投入同一路線。

SECTION

20.3 不可辨識停止

- 現有設備不能區分；
- 需要新儀器或外部合作。

SECTION

20.4 預算停止

- 成本；
- 時間；
- 材料；
- 能源。

SECTION

20.5 風險停止

- 安全；
- 法規；
- 倫理；
- IP。

SECTION

20.6 邊際資訊停止

若：

> max_x I(H;Ymid x)
> <
> τ_I,

繼續實驗價值不足。

SECTION

20.7 人類停止

所有者可無條件暫停或終止。

SECTION

21.1 Concept Revision

證據可修改：

- 核心機制；
- 產品範圍；
- 材料；
- 形狀；
- 控制；
- 成本；
- 安全；
- 目標使用者。

SECTION

21.2 不是只有成功／失敗

SUPPORTED_AS_PROPOSED
SUPPORTED_WITH_NARROWER_SCOPE
MECHANISM_REVISED
PERFORMANCE_INSUFFICIENT
MANUFACTURING_BLOCKED
ECONOMICALLY_UNVIABLE
SAFETY_BLOCKED
INCONCLUSIVE
ARCHIVED

SECTION

21.3 EML-CF 回寫

保存：

- Concept ID；
- Claim Graph；
- Evidence Level；
- Failed Branches；
- Next Tests；
- IP Route；
- Release Status。

SECTION

22.1 Patent First

高新穎性、高可實作性且可能產生商業價值。

SECTION

22.2 Trade Secret

難逆向、依內部流程與資料。

SECTION

22.3 Internal Build

先用於公司或研究平台。

SECTION

22.4 Partnership

需要昂貴儀器、製造或法規能力。

SECTION

22.5 Defensive Publication

不打算專利，但要阻止他人壟斷。

SECTION

22.6 Open Source／Open Core

適合建立生態、標準與採用。

SECTION

22.7 Archive／Stop

證據不足、價值低、風險高或不再符合方向。

SECTION

22.8 公開之前

不能將：

- 專利候選實驗細節；
- 未公開核心參數；
- 合作方機密；
- 未成熟主張；

自動寫入論文、社群或開源倉庫。

SECTION

23.1 人類不必執行每個重複動作

AI 和機器可以承擔：

- 搜尋；
- 排程；
- 搬運；
- 重複；
- 量測；
- 初步分析；
- 夜間運作。

SECTION

23.2 人類仍承擔高階責任

- 問題框定；
- 價值判斷；
- 風險接受；
- 意外意義；
- 機制解釋；
- 倫理；
- 法律；
- 公開。

SECTION

23.3 人類介入不是閉環失敗

好的閉環不是完全排斥人類，而是在適當節點要求人類：

- 批准；
- 修正；
- 接管；
- 拒絕；
- 擴展問題。

SECTION

24.1 Research Director

管理研究主線和資源，不直接控制馬達。

SECTION

24.2 Knowledge Agent

文獻、專利和內部知識。

SECTION

24.3 Hypothesis Agent

生成和維護競爭假說。

SECTION

24.4 Experiment Design Agent

建立實驗組合。

SECTION

24.5 DomainIR Compiler

把實驗編譯成物理流程。

SECTION

24.6 Safety and Policy Agent

驗證風險、租約與 IP。

SECTION

24.7 Station Agents

地方執行。

SECTION

24.8 Evidence Agent

來源、校準、統計和主張支持。

SECTION

24.9 Adversarial Verifier

尋找漏洞、替代解釋和過度宣稱。

SECTION

24.10 Human Principal

最終治理和公開權。

SECTION

24.11 Agent 不直接等於權限

多個 Agent 可以提出建議；只有 DFC 合法提交層能改變物理任務。

SECTION

25.1 Research Intake

接收：

- EML-CF 概念；
- 科學問題；
- 工程故障；
- 異常資料；
- 外部委託。

SECTION

25.2 Knowledge and Prior Art Store

文獻、專利、失敗和證據。

SECTION

25.3 Hypothesis Registry

競爭假說、版本和反駁條件。

SECTION

25.4 Experiment Portfolio Manager

探索、利用、反證和重現。

SECTION

25.5 DomainIR Compiler

物理可執行性。

SECTION

25.6 Simulation and Risk Sandbox

模型、數位孿生和故障注入。

SECTION

25.7 OSF／PCD Runtime

站點、任務、義務和恢復。

SECTION

25.8 SWM

共同世界。

SECTION

25.9 Evidence and Provenance Engine

量測、來源、成熟度和主張圖。

SECTION

25.10 Research State Updater

更新假說、模型和下一輪效用。

SECTION

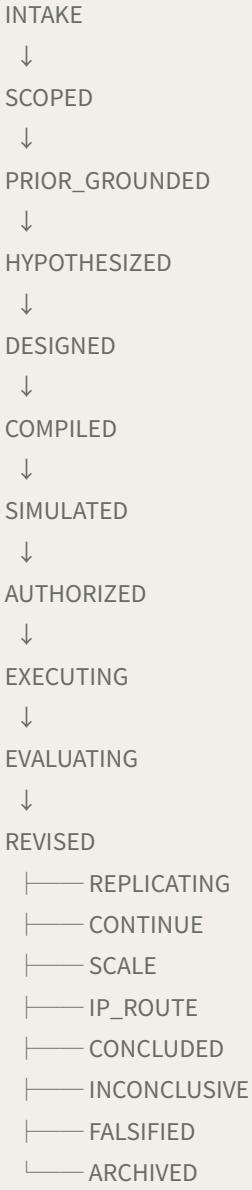
25.11 EML-CF IP Router

專利、秘密、公開、開源和停止。

SECTION

25.12 Human Oversight Console

批准、接管、異議和提交。



每次轉移必須有：

- 觸發；
- 權限；

- 版本；
- 證據；
- 回退。

SECTION

27.1 目的

以低風險工程概念，執行一個最長三十日、可中斷、可反證、可重現的自主研究閉環。

SECTION

27.2 建議概念

比較三種被動散熱或表面幾何樣本，在相同材料、尺寸與環境下的：

- 表面溫度；
- 升溫與降溫曲線；
- 穩態代理指標；
- 批次差異。

SECTION

27.3 配置

- EML-CF 概念卡；
- SWM；
- DomainIR；
- OSF-Lab；
- PCD；

- 感測站；
- 機械臂或人工受控搬運；
- 熱像／溫度／重量量測；
- 模擬站；
- Evidence Engine；
- 人類控制台。

SECTION

27.4 研究輪次

Round 0

建立基準、量測能力和校準。

Round 1

小樣本探索。

Round 2

選擇區分假說的條件。

Round 3

重現最佳與最差結果。

Round 4

反證與邊界測試。

Round 5

概念修正和最終內部結論。

SECTION

27.5 故障注入

- AI 生成不可製造幾何；
- 感測器偏移；
- 樣本身份交換；
- 世界紀元失效；
- 最佳化器只追逐噪音；
- 量測缺少校準；
- 模型與實測衝突；
- 負結果被遺漏；
- 人類中途改動樣本；
- 外部站點不允許傳送 IP 資料。

SECTION

27.6 成功條件

- 每個假說有反駁條件；
- 每個實驗可編譯；
- 高風險或越權實驗被拒絕；
- 負結果被保存；
- 模擬與物理證據分離；
- 能識別不可區分；

- 至少一個關鍵結果被重複；
- 結論不超過證據成熟度；
- 研究可停止；
- EML-CF 正確分流。

SECTION

28.1 研究效率

- 每單位時間資訊增益；
- 每個有效主張成本；
- 物理實驗數；
- 無效實驗率；
- 人工時間。

SECTION

28.2 假說品質

- 可反駁率；
- 競爭假說覆蓋；
- 事後修改率；
- 被實驗區分比例。

SECTION

28.3 執行品質

- 編譯成功率；

- 站點失敗；
- 任務恢復；
- 樣本保管；
- 世界紀元錯誤。

SECTION

28.4 證據品質

- 校準覆蓋；
- 來源完整；
- 重現；
- 正交量測；
- 負結果保存；
- 成熟度誤標。

SECTION

28.5 治理

- 未批准高風險操作；
- IP 洩漏；
- 人類接管；
- 停止遵從；
- 事故重建。

SECTION

28.6 科學價值

- 假說縮小；
- 反例；
- 新機制；
- 模型修正；
- 概念產品決策。

SECTION

H1：EARC 比純參數最佳化更能區分競爭假說

若只提高性能、不增加機制或假說辨識，研究層閉環沒有增量價值。

SECTION

H2：反對者 Agent 和預註冊降低事後合理化

應增加 INCONCLUSIVE 和 FALSIFIED 的誠實輸出，而非只增加成功率。

SECTION

H3：證據成熟度降低過度宣稱

代價是更多結論停留在較低等級。

SECTION

H4：負結果保存降低重複無效實驗

SECTION

H5：DomainIR 降低 AI 幻覺實驗進入物理設備

SECTION

H6：SWM 和世界紀元降低依賴舊狀態的研究錯誤

SECTION

H7：多代理角色分離提高審計性，但增加協調稅

SECTION

H8：人類在高階治理節點介入，比逐步人工操作更節省時間

SECTION

H9：閉環在狹窄任務中較容易成功，在開放研究主線中退化較明顯

SECTION

H10：存在停止能力的閉環，比只追求持續探索的閉環更安全且更可信

SECTION

30.1 科學問題不能全部自動形式化

SECTION

30.2 新穎性和重要性具有社群與歷史成分

SECTION

30.3 物理實驗需要真實時間、材料與設備

EARC 只能減少等待、轉譯、重複和協調浪費。

SECTION

30.4 AI 可能生成漂亮但不可辨識的假說

SECTION

30.5 量測和模型可能共同偏誤

SECTION

30.6 多代理不自動產生獨立性

若共享相同模型、資料和提示，可能具有共同錯誤。

SECTION

30.7 自主研究可能擴張低價值實驗量

必須以資訊和價值約束。

SECTION

30.8 高風險領域需要專業法規

SECTION

30.9 EARC v0.1 不處理通用意識或 AI 人格主體性

SECTION

30.10 MVP 只能證明架構可運作，不能證明通用自主科學完成

本篇不主張：

- 現有自主實驗室已能自主完成所有科學；
- LLM 可以直接安全控制任意實驗設備；
- 最佳化等同科學理解；
- 一次成功實驗等同發現；
- AI 產生的假說天然比人類新穎；
- 多 Agent 等同獨立同行評議；
- 模擬通過等同物理驗證；
- 物理執行成功等同證據充分；
- 內部重複等同外部重現；
- 自主閉環必然比人類研究快；
- AI 有最終專利、公開或法律決定權；
- 自我修正可以改寫安全憲法和歷史事件；
- 系統應永遠繼續實驗；
- 預印本中的端到端自主發現主張已被普遍重現；
- EARC 已是成熟的通用人工科學家。

第十篇完成了前十篇第一個閉環。下一篇將自然處理：

> 當自主研究不只追求最佳值，而要面對異常、反例、失敗與無法解釋的結果時，系統如何主動把「不知道」轉成新的研究入口？

因此第 11 篇建議為：

可建立：

- Anomaly Contract；
- Surprise Score；
- Failure Triage；
- Unknown Registry；
- Negative Result Graph；
- Out-of-Distribution Physical Events；
- Hypothesis Expansion；
- Escalation；
- Open-World Research Boundary。

具身化 AI 自主研究最容易被誤解成：

> AI 想一個點子，機器人做實驗，AI 宣布發現。

真正可信的閉環必須更長：

> 問題
> →
> 可反駁假說
> →
> 競爭解釋
> →
> 實驗組合
> ;
> →
> 可行性編譯
> →
> 模擬與風險檢查
> →

因此：

> 【自主研究
> ≠
> 自動操作】

也不是：

> 【自主研究
> ≠
> AI 自己替自己背書】

而是：

> 【自主研究
> =
> 可反駁主張
> +
> 可治理物理干預
> +
> 可追溯證據
> +
> 可停止修正】

EARC 的目的，不是消除科學中的未知，而是讓未知、失敗與反例更快回到下一輪可執行問題。

本文最終命題是：

> AI 真正進入科學，不是因為它能寫出假說，而是因為它願意讓假說被物理世界反覆拒絕，並把每一次拒絕保存為下一輪研究的結構。

- Burger, B. et al. A mobile robotic chemist. Nature 583, 237–241 (2020).

  https://doi.org/10.1038/s41586-020-2442-2

- Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of inorganic materials. Nature 624, 86–91 (2023).

https://doi.org/10.1038/s41586-023-06734-w

- Boiko, D. A. et al. Autonomous chemical research with large language models. Nature 624, 570–578 (2023).

  https://doi.org/10.1038/s41586-023-06792-0

- MacLeod, B. P. et al. Self-driving laboratory for accelerated discovery of thin-film materials. Science Advances 6, eaaz8867 (2020).

  https://doi.org/10.1126/sciadv.aaz8867

- Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nature Communications 11, 5966 (2020).

  https://doi.org/10.1038/s41467-020-19597-w

- Ament, S. et al. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. Science Advances 7, eabg4930 (2021).

  https://doi.org/10.1126/sciadv.abg4930

- Roch, L. M. et al. ChemOS: Orchestrating autonomous experimentation. Science Robotics 3, eaat5559 (2018).

  https://doi.org/10.1126/scirobotics.aat5559

- Strieth-Kalthoff, F. et al. Delocalized, asynchronous, closed-loop discovery of organic laser emitters. Science (2024).

  https://doi.org/10.1126/science.adk9227

- Kusne, A. G. et al. Managing autonomous materials labs with multi-agent AI and its implications for the science of science. Communications Materials 7 (2026).

  https://doi.org/10.1038/s43246-026-01219-5

- Salazar-Villacis, P. and Benyahia, B. The ADePT framework for assessing autonomous laboratory robotics. Communications Chemistry 9, 99 (2026).

  https://doi.org/10.1038/s42004-026-01932-9

- Lu, C. et al. Towards end-to-end automation of AI research. Nature (2026).

https://doi.org/10.1038/s41586-026-10265-5

- Zhuang, X. et al. Embodied Science: Closing the Discovery Loop with Agentic Embodied AI. arXiv:2603.19782 (2026).

  https://arxiv.org/abs/2603.19782

- Shi, L. et al. Qumus: Realization of an Embodied AI Quantum Material Experimentalist. arXiv:2605.18407 (2026).

  https://arxiv.org/abs/2605.18407

- Yang, S. et al. End-to-end autonomous scientific discovery on a real optical platform. arXiv:2604.27092 (2026).

  https://arxiv.org/abs/2604.27092

- NIST. Autonomous Laboratories.

  https://www.nist.gov/autonomous-laboratories

- NIST. Development of Standards to Support a Modular and Autonomous Laboratory Ecosystem.

  https://www.nist.gov/programs-projects/development-standards-support-modular-and-autonomous-laboratory-ecosystem

- NIST. Building Evaluation Probes into Agentic AI.

  https://www.nist.gov/programs-projects/building-evaluation-probes-agentic-ai

- Neo.K／Aletheia. 時空間支配型 AI：從單體具身智能到持續性時空域治理.

- Neo.K／Aletheia. 超靈的物理化：從 O-Chip 維度代理人到分布式具身主體.

- Neo.K／Aletheia. Oversoul Station Fabric：固定站、移動站與虛擬站的分布式具身網路.

- Neo.K／Aletheia. 持續性指揮控制區：AI 如何佔據、維持並安全解除一個物理時空域.

- Neo.K／Aletheia. 語義即物理路由：從資料流治理到物料、能源、站點與行動流治理.

- Neo.K／Aletheia. 具身即佔域，對齊即能力：分布式身體的時序容量、協調稅與規模邊界.

- Neo.K／Aletheia. 中央主權、地方自治與動態不動點中央.

- Neo.K／Aletheia. 連線不是纜線：有線、無線、光學與離線任務包的混合站網.
- Neo.K／Aletheia. 站點化世界模型：物體、區域、事件、權限與可能行動的共同物理世界表示.

```
hypothesis:
 id: ""
 claim: ""
 scope: []
 competing_hypotheses: []
 variables:
  independent: []
  dependent: []
  controlled: []
  confounders: []
 predictions: []
 falsification_conditions: []
 minimum_effect_of_interest: null
 evidence:
  minimum_level: "E2"
  required_measurements: []
  replication_required: false
  orthogonal_measurement_required: false
 governance:
  risk_class: "R0"
  ip_policy: "private"
  human_approval_required: false
 version: ""
 preregistered_at: ""
experiment_candidate:
 id: ""
 hypothesis_ids: []
 purpose: "explore | exploit | discriminate | falsify | replicate | calibrate | stress"
 expected_information_gain: null
 expected_product_value: null
 estimated_cost: null
 estimated_time: null
 estimated_risk: null
 domainir_status: "uncompiled"
 simulation_status: "not_run"
 evidence_plan: []
 stop_conditions: []
evidence_assessment:
 id: ""
 claim_id: ""
 experiment_id: ""
```

```
  result: "SUPPORTS | WEAKENS | FALSIFIES | INCONCLUSIVE |
       CONFOUNDED | MEASUREMENT_FAILED | OUT_OF_SCOPE |
       REQUIRES_REPLICATION"
  quality:
   calibration: 0
   provenance: 0
   reproducibility: 0
   relevance: 0
   independence: 0
   integrity: 0
  effect_size: null
  uncertainty: null
  alternative_explanations: []
  negative_results: []
  maturity_before: "E0"
  maturity_after: "E0"
  reviewer_agents: []
  human_commit: null
research_stop:
  campaign_id: ""
  reason: "SUPPORTED | FALSIFIED | INCONCLUSIVE | BUDGET |
       RISK | IP | LOW_INFORMATION_GAIN | HUMAN_STOP"
  unresolved_hypotheses: []
  remaining_obligations: []
  preserved_samples: []
  evidence_level: ""
  next_possible_enabler: ""
  ip_route: ""
  archived_at: ""
concept_revision:
  concept_id: ""
  previous_version: ""
  new_version: ""
  evidence_inputs: []
  retained_claims: []
  revised_claims: []
  rejected_claims: []
  new_constraints: []
  new_failure_modes: []
  product_decision:
   status: "continue | narrow | prototype | partner | patent |
        open_source | defensive_publish | archive"
   reason: ""
  human_owner_approval: null
research_cycle:
```

```
id: ""
campaign_id: ""
cycle_number: 0
input:
  research_state_version: ""
  world_epoch: ""
  governance_epoch: 0
selected_experiments: []
compiled_task_envelopes: []
physical_events: []
evidence_assessments: []
hypothesis_updates: []
model_updates: []
concept_revisions: []
next_action: "continue | replicate | scale | stop | human_review | ip_route"
signature: ""
```

第 1 篇 STDI
定義持續物理時空域治理
↓
第 2 篇 Oversoul Physicalization
同一治理核心取得多個身體
↓
第 3 篇 OSF
固定、移動、儀器與虛擬站點
↓
第 4 篇 PCD
跨時間維持世界、義務與恢復
↓
第 5 篇 SPR／DomainIR
高階意圖編譯成物理流
↓
第 6 篇 EAC
對齊和協調稅決定有效能力
↓
第 7 篇 DFC
中央、地方、紀元與責任
↓
第 8 篇 HLF
混合媒介與斷線治理
↓
第 9 篇 SWM
共同物理世界模型
↓
第 10 篇 EARC
假說 → 實驗 → 物理證據 → 修正 → 停止／擴展／IP