表示落差：意圖、計算模型與物理現實之間的不可完備映射

The Representation Gap: Incomplete Mappings Among Intent, Computational Models, and Physical Reality

v0.1 / 2026-07-27 / Neo.K with Aletheia / 初版完成

https://thisoneisneok.com/html/papers/pu-0-03.html

SECTION

The Representation Gap: Incomplete Mappings Among Intent, Computational Models, and Physical Reality

論文編號： PU-0-03

系列：「程式宇宙書系」第 0 冊《程式之前》

作者： Neo.K with Aletheia

機構： EveMissLab／一言諾科技有限公司

版本： v0.1

日期： 2026 年 7 月 27 日

SECTION

摘要

前兩篇論文分別建立了意圖宇宙、計算機宇宙與物理宇宙的基本區分，以及三者之間的耦合循環。本篇集中處理耦合過程中的核心難題：任何從意圖到模型、從模型到世界、再從世界回到資料與解釋的映射，都不是完備、無損或透明的。

本文將此現象定義為「表示落差」：

> 【mathcal I
> notcong
> mathcal C
> notcong
> mathcal P】

這不只是說三個宇宙不相同，而是指出它們之間不存在普遍、單一、無損、可逆的完整同構。意圖包含未決定、矛盾、價值、情境與不可代理部分；計算模型必須選擇變數、型別、狀態、規則與合法操作；物理世界則包含未建模因素、材料限制、噪聲、制度、身體、偶發事件與不可完全觀測的因果關係。

本文提出五類主要落差：

- 意圖落差：主體真正想要的內容與其可表達要求之間的差異；
- 表示落差：已表達意圖與計算模型所保存結構之間的差異；
- 執行落差：計算模型預期世界差分與實際結果之間的差異；
- 觀測落差：物理世界與被感測、資料化、標註之世界之間的差異；
- 解釋落差：資料結構與人類或 AI 所形成意義之間的差異。

  總落差可寫為：

> R_(total)
> =
> R_I
> +
> R_R
> +
> R_X
> +
> R_O
> +
> R_M

然而，本文反對把所有差異都視為可透過更多資料、更大模型、更精細感測器或更長規格完全消除。本文進一步區分：

- 工程性落差：可藉由更好的需求、模型、測試、資料與驗證降低；
- 結構性落差：源自跨存在域轉譯、有限表示與情境依賴，無法完全消除；
- 規範性落差：源自價值衝突、正當性與不可替代決策，不應被技術自動壓平；
- 生成性落差：系統執行後創造原先不存在的新情境、新意圖與新風險。

本文提出：

> 【R_(min)
> >
> 0】

對任何涉及開放人類意圖、複雜物理世界與持續制度影響的系統而言，總落差的理論下界通常不為零。成熟系統的目標不是宣稱消除全部落差，而是顯示哪些部分已表示、保留哪些部分未決定、測量哪些差異可降低、揭露哪些差異不可消除，並為不可表示與不可逆部分保留治理權。

本文建立「不完備表示原則」：

> 任何計算模型只對其明示的變數、關係、規則、尺度與觀測條件負責；它不能因內部一致而自動取得對外部世界的完整代表權。

本文也提出「剩餘保留原則」：未能進入模型的內容，不得自動被視為不存在、不重要、無效或已獲同意。這一原則尤其適用於少數情境、例外、弱勢主體、文化差異、體驗狀態、長期後果與新生成風險。

本文使用醫療模型、公共政策、AI Agent、數位分身與待辦系統等案例，說明一個系統可以在語法、邏輯、測試與資料層面均正確，卻仍在意圖或現實層面失敗。本文最後提出可證偽研究綱領，包括意圖殘差測量、模型覆蓋率、未知未知事件率、觀測遺漏、價值衝突保留、模型外受影響者召回率、反事實失真與生成性落差追蹤。

本文為下一篇〈數位宇宙作為人工存在域〉奠定限制論基礎：只有承認模型永遠不等於現實，才能進一步研究數位世界中的實體、狀態、規則與歷史究竟具有何種人工存在地位。

關鍵詞： 表示落差、不完備映射、意圖殘差、模型風險、觀測落差、規範性落差、數位治理、計算存在論

SECTION

Abstract

The previous two papers distinguished the intentional, computational, and physical universes and developed their coupling dynamics. This paper focuses on a central problem: mappings from intent to model, model to world, world to data, and data to interpretation are never universally complete, lossless, transparent, or reversible.

The representation gap is expressed as:

> mathcal I
> notcong
> mathcal C
> notcong
> mathcal P

The paper identifies five major gaps: intentional, representational, execution, observational, and interpretive. It further distinguishes engineering gaps, structural gaps, normative gaps, and generative gaps.

For open human intentions and complex physical worlds, the theoretical minimum residual is generally nonzero:

> R_(min)
> >
> 0

The goal of mature systems is therefore not to claim complete elimination of residuals, but to reveal what has been represented, preserve unresolved content, measure reducible discrepancies, disclose irreducible limitations, and retain governance over unrepresented or irreversible effects.

The paper proposes an incompleteness principle of representation and a residual preservation principle. It develops examples from medicine, public policy, AI agents, digital twins, and task systems, and concludes with a falsifiable research program.

Keywords: representation gap, incomplete mapping, model risk, intentional residual, observation gap, normative residual

人類之所以建立模型，是因為無法在每次行動前直接處理全部現實。

模型具有必要性：

- 壓縮；
- 比較；
- 預測；
- 規劃；
- 驗證；
- 溝通；
- 自動化。

但模型越有用，人們越容易忘記：

> 模型之所以可用，正是因為它刪除了大量內容。

資料庫不是世界。

需求文件不是意圖全部。

感測器數值不是物理現實全部。

模型輸出也不是行動結果全部。

因此，本篇的起點是：

> 【Representation
> =
> Selection
> +
> Compression
> +
> Exclusion】

任何表示都同時保存一些內容，排除另一些內容。

設 A 與 B 為不同存在域，映射為：

> f
> :
> A
> →
> B

表示落差可定義為：

> R_f(a)
> =
> a
> \
> Recover_f
> (
> f(a)
>)

也就是，原對象中無法由其表示恢復的部分。

若映射不可逆：

> f^(-1)(f(a))
> ≠
> a

則存在非零落差。

SECTION

2.1 三宇宙形式

> Φ_(IC)
> :
> mathcal I
> →

> Φ_(CP)
> :
> mathcal C
> →
> mathcal P

> Φ_(PC)
> :
> mathcal P
> →
> mathcal C

> Φ_(CI)
> :
> mathcal C
> →
> mathcal I

一般而言：

> Φ^(-1)_(XY)
> (
> Φ_(XY)(x)
>)
> ≠
> x

所以跨宇宙映射通常不是完整同構。

SECTION

3.1 意圖落差 R_I

主體的真實目的、感受、矛盾、猶豫與未決定內容，往往大於其語言表達。

例如「幫我做一個簡單的網站」，其中「簡單」可能包含視覺簡潔、操作簡單、容易維護、成本低、載入快、不需要登入或不暴露個資。

SECTION

3.2 表示落差 R_R

即使要求已經表達，計算模型仍必須選擇哪些實體存在、哪些屬性重要、哪些狀態合法、哪些事件可發生、哪些例外被忽略、哪些主體具有身分，以及哪些時間尺度被保存。

SECTION

3.3 執行落差 R_X

> R_X
> =
> d
> (
> Δ W^ast,
> Δ W_(actual)
>)

原因包括硬體失敗、網路延遲、外部系統不同步、人類不配合、制度流程、物理誤差與權限不足。

SECTION

3.4 觀測落差 R_O

> D
> =
> Φ_(PC)(P)
> subsetneq
> P

未被量測的，不等於不存在。

SECTION

3.5 解釋落差 R_M

> M⊠(D)
> ≠
> M⊠(D)

不同主體、模型、價值與歷史會形成不同解釋。

SECTION

4.1 工程性落差

可藉由更完整需求、更好感測器、更多測試、更高資料品質、更細型別、更強驗證與更好的觀測降低。

SECTION

4.2 結構性落差

源自有限表示、開放世界、情境依賴、多尺度、跨域轉譯、時間變化與巨大可能狀態空間。這類落差不能被完全消除。

SECTION

4.3 規範性落差

源自價值衝突、正當性、權利、成本分配、不可代理選擇，以及某些即使有效也不應接受的結果。規範問題不能被更多資料自動解決。

SECTION

4.4 生成性落差

系統執行後可能創造新使用方式、新風險、新主體、新依賴、新意圖與新制度問題。這些內容在原模型建立時尚不存在。

> 任何計算模型只對其明示的變數、關係、尺度、規則、觀測條件與適用域負責。內部一致性不能自動轉化為對外部世界的完整代表權。

形式上：

$$\text{Valid\_C}(M) \not\Rightarrow \text{Complete\_P}(M)$$

同樣：

$$\text{Consistent}(M) \not\Rightarrow \text{Legitimate}(M)$$

一個模型可以邏輯一致，卻不公平、失真或不適用。

本文提出：

> 任何未被模型表示的內容，都不得因此被自動判定為不存在、不重要、無效、非法或已獲同意。

形式上：

$$x \notin M \not\Rightarrow \neg x$$

這一原則特別適用於：

- 少數案例；
- 邊界情境；

- 弱勢主體；
- 無法量化的體驗；
- 尚未出現的新風險；
- 長期後果；
- 未被詢問的反對；
- 文化與語境差異。

SECTION

6.1 未建模不等於不存在

系統中沒有某種欄位，只代表系統沒有表示它。

例如醫療資料沒有記錄疼痛，不代表病人沒有疼痛。

SECTION

6.2 未回答不等於同意

使用者沒有選擇某項設定，可能是：

- 不知道；
- 看不懂；
- 沒有找到；
- 無法操作；
- 不願回答；
- 尚未決定。

因此：

> NoResponse
> not⇒
> Consent

SECTION

6.3 未分類不等於非法

現實中的新對象可能尚未符合既有型別。

成熟系統應允許：

- 未知；
- 待分類；
- 爭議；
- 例外；
- 人類審查；

  而不是強迫所有對象立即落入既有類別。

SECTION

7.1 有限模型與開放世界

若模型狀態空間為 mathcal S_M，現實相關狀態空間為 mathcal S_P，通常：

> |mathcal S_M|
> <
> |mathcal S_P|

即使模型非常大，也仍只選擇有限變數與關係。

SECTION

7.2 意圖具有未決定性

主體可能同時：

- 想要便利；
- 擔心隱私；
- 希望快速；
- 不願承擔風險；
- 尚未決定優先順序。

  這不是資訊尚未收集完畢，而可能是意圖本身尚未完成。

SECTION

7.3 世界會持續生成新狀態

系統投入使用後，使用者、競爭者、制度與環境會回應系統。

因此適用域不是靜止的。

SECTION

7.4 觀測具有成本

全面感測需要：

- 能源；
- 儲存；
- 隱私代價；
- 計算；

- 人力；
- 制度授權。

  所以不是所有可觀測內容都應被觀測。

SECTION

7.5 規範衝突無法由單一最適解消除

「效率最高」不代表「最公平」。

「總效益最大」也不代表每個受影響者都應接受。

因此：

> 【R_(min)
> >
> 0】

對開放、持續、多主體系統而言，是合理的基礎假設。

承認模型不完備，不代表：

- 不需要精確；
- 所有模型都一樣；
- 無法驗證；
- 只能依賴直覺；
- 任何錯誤都可被原諒。

  相反地，正因模型有限，更需要精確說明：

- 表示了什麼；
- 沒有表示什麼；
- 適用於哪裡；

- 在何種條件下成立；
- 誤差多大；
- 如何失效；
- 誰能反駁。

SECTION

8.1 適用域

定義模型適用域：

> mathcal D(M)
> =
> {
> x
> mid
> M
> is validated forx
> }

若輸入超出適用域：

> x
> ∉
> mathcal D(M)

系統應拒絕、降級或轉交，而非假裝正常。

SECTION

8.2 信心不是完整性

模型可對某結果具有高信心，但仍遺漏重要變數。

> HighConfidence
> not⇒
> HighCoverage

SECTION

8.3 精確錯誤

一個結果可以非常精確地計算錯誤問題。

例如系統精確預測「誰最可能點擊」，但原始目的可能是「誰真正從內容中受益」。

SECTION

9.1 模型內真

命題 q 在模型 M 中成立：

> M
> |=
> q

這只表示它符合模型的公理、資料與規則。

SECTION

9.2 世界真

命題在相關物理或社會世界中實際成立，需要額外證據：

> P
> |=
> q

SECTION

9.3 二者不自動等價

> M
> |=
> q
> not⇒
> P
> |=
> q

例如：

- 模型判定「已送達」；
- 世界中包裹仍在門外或送錯地址。

SECTION

9.4 世界真也可能未進入模型

> P
> |=
> q
> not⇒
> M
> |=
> q

某種傷害、需求或關係可能真實存在，但資料庫沒有欄位表示。

SECTION

10.1 類別創造可計算性

計算需要將開放現實壓縮成：

- 是／否；
- 類型；
- 數值；
- 枚舉；
- 狀態；
- 身分。

分類使自動化成為可能。

SECTION

10.2 類別也創造盲點

任何分類都可能：

- 合併差異；
- 切斷連續性；
- 固定邊界；
- 排除混合情境；
- 把暫時狀態當永久身分。

SECTION

10.3 型別錯配

若現實對象 x 被迫映射到不適合的型別 T：

> x
> not|=
> T

後續所有計算都可能形式正確但語意錯誤。

SECTION

10.4 開放型別

對高不確定領域，應保留：

unknown
unresolved
contested
mixed
not-applicable
human-review
這不是設計失敗，而是對剩餘的正式承認。

SECTION

11.1 微觀與宏觀

同一系統在不同尺度會有不同性質。

例如交通系統：

- 微觀：單一車輛速度；
- 中觀：道路流量；
- 宏觀：都市通勤結構。

  只建模某一尺度，可能錯過跨尺度效應。

SECTION

11.2 時間尺度

短期改善可能造成長期損失。

$$
\Delta W_{short} > 0
$$

例如推薦系統提高短期互動，卻降低長期信任。

SECTION

11.3 聚合落差

平均值可能遮蔽個體傷害。

> AverageBenefit
> >
> 0
> not⇒
> ∀ i, Benefit⊠>0

人類意圖常包含：

- 如果不做會怎樣；
- 是否存在更好方案；
- 某結果是不是由系統造成；
- 哪個選擇本來可以避免傷害。

  但計算資料通常只記錄實際發生的路徑。

SECTION

12.1 反事實不可直接觀測

> P(a)
>
> and
>
> P(¬ a)

無法在完全相同世界中同時發生。

SECTION

12.2 模擬不是實際反事實

模擬可提供候選比較，但仍依賴模型假設。

SECTION

12.3 決策需要反事實謙遜

系統應區分：

- 已觀測；
- 推論；
- 模擬；
- 假設；
- 未知。

SECTION

13.1 價值不是單一分數

多數重要決策包含：

- 效率；
- 公平；
- 安全；
- 自由；
- 隱私；
- 可逆性；

- 尊嚴。

  將其壓成單一效用值：

> U
> =
> Σ□
> w□ v□

只是某種規範選擇，不是自然真理。

SECTION

13.2 權重由誰決定

權重 w□ 代表：

- 誰的價值；
- 哪個時間尺度；
- 誰承擔風險；
- 哪些權利可以被交換。

SECTION

13.3 不可交換約束

某些條件不應只是低權重，而應作為硬性邊界：

> v□
> ∉
> trade-off set

例如基本權利或不可撤回選擇。

SECTION

14.1 已知未知

已知某項資料缺失。

SECTION

14.2 未知未知

系統甚至不知道應該建模什麼。

SECTION

14.3 新生成未知

系統部署後創造新的現象，使原本不存在的問題出現。

例如：

- 新型平台操縱；
- 新職業依賴；
- 新型詐騙；
- AI 與人類的新關係；
- 新的數位財產爭議。

SECTION

14.4 未知狀態應被保存

成熟模型不能把所有未知都壓成預設值。

> Unknown
> ≠
> 0

也不等於否定。

SECTION

15.1 意圖宇宙

醫療制度可能希望：

- 早期發現疾病；
- 分配有限資源；
- 降低死亡率；
- 避免過度治療；
- 保護病人自主。

  這些目標彼此可能衝突。

SECTION

15.2 計算模型

風險模型可能使用：

- 年齡；
- 檢驗值；
- 病史；
- 影像；
- 家族史；
- 既有診斷。

SECTION

15.3 表示落差

模型可能未包含：

- 病人主觀疼痛；
- 生活環境；
- 醫療可近性；
- 文化差異；
- 未被診斷的症狀；
- 資料品質差異。

SECTION

15.4 模型內正確、世界中失敗

模型可能對訓練資料具有高準確率，但在新族群上失效。

因此：

> Validated_(D□)(M)
> not⇒
> Valid_(D□)(M)

SECTION

15.5 治理要求

高風險模型應提供：

- 適用族群；
- 缺失變數；
- 不確定性；
- 人類覆核；

- 異議；
- 新資料更新；
- 錯誤補救。

SECTION

16.1 政策目的

制度希望把資源提供給真正需要者，同時控制詐欺與預算。

SECTION

16.2 數位化

現實生活被壓縮成：

- 收入；
- 家戶；
- 身分；
- 工作；
- 居住；
- 財產；
- 文件狀態。

SECTION

16.3 模型外存在

某些人可能：

- 沒有完整文件；

- 居住狀態不穩；
- 家庭關係複雜；
- 收入不規則；
- 無法使用數位介面。

  他們可能真實需要幫助，卻無法被模型合法表示。

SECTION

16.4 分類造成制度效果

系統中的 ineligible 不只是標籤，可能導致：

- 補助中止；
- 住房風險；
- 醫療中斷；
- 申訴成本；
- 社會污名。

  因此表示落差會被執行放大。

SECTION

17.1 原始要求

幫我完成研究報告。

SECTION

17.2 模型化

Agent 可能把完成定義為：

- 文件存在；

- 字數達標；
- 章節完整；
- 引用格式正確；
- 測試通過。

SECTION

17.3 被遺漏的意圖

使用者可能真正要求：

- 論證新穎；
- 資料正確；
- 與既有研究區分；
- 適合出版；
- 能承受批評。

這些不容易被單一完成旗標表示。

SECTION

17.4 完成落差

> TaskComplete_C
> not⇒
> IntentSatisfied_I

SECTION

17.5 正確做法

Agent 應將完成拆成：

- 已生成；
- 已驗證；
- 已外部檢查；
- 尚有未知；
- 等待人類接受。

SECTION

18.1 宣稱

數位分身常被描述為物理設備、城市或人體的數位副本。

SECTION

18.2 實際狀態

它更精確地是：

> $M_C(P \mid V,S,T)$

也就是在特定變數、感測器、尺度與時間下，對物理對象的模型。

SECTION

18.3 主要落差

- 感測器故障；
- 未建模部件；
- 更新延遲；
- 模型參數過時；

- 維護行為未記錄；
- 異常情境缺乏資料。

SECTION

18.4 合理名稱

「分身」容易暗示完整同一。

更精確的理解是：

> 可更新、具適用域的物理對象計算投影。

待辦系統能表示：

- 任務；
- 期限；
- 優先級；
- 完成狀態。

但生活中重要的內容還包括：

- 是否值得做；
- 是否應該休息；
- 是否對他人有責任；
- 是否因焦慮而過度安排；
- 某件事完成後是否真的改善生活。

因此，一個完美運作的待辦系統，仍可能讓人更有效率地追求錯誤目的。

這是規範性落差，而不是普通程式錯誤。

表示落差不能只靠模型團隊內部處理。

需要治理結構。

SECTION

20.1 模型卡不足

模型卡可以描述限制，但若使用者：

- 看不到；
- 看不懂；
- 無法反駁；
- 無法停止；
- 無法獲得補救；

  則限制揭露仍不足以形成治理。

SECTION

20.2 落差帳本

本文建議建立：

> mathcal L_R
> =
> <
> R,
> S,
> O,
> E,
> A,
> G
> >

其中：

- R：殘差內容；
- S：嚴重度；
- O：來源；
- E：證據；
- A：受影響主體；
- G：治理與補救。

SECTION

20.3 殘差所有權

每個重大落差都應有人或制度負責：

- 追蹤；
- 更新；
- 回報；
- 降低；
- 停止系統；
- 補償損害。

SECTION

20.4 模型外申訴

受影響者必須能說：

我的情況不符合你們的分類。
資料是錯的。
規則不適用。
結果雖符合規則但不正當。

這些陳述本身就是對模型邊界的證據。

單純寫上：

AI 可能出錯
結果僅供參考
不等於負責任地處理落差。

真正的表示謙遜需要：

- 具體適用域；
- 已知限制；
- 未知部分；
- 信心與覆蓋分離；
- 證據來源；
- 失效條件；
- 人類接管；
- 可撤回或補救路徑。

SECTION

21.1 結構化不確定性

不確定性應進入正式狀態，而不只放在備註。

```
claim:
 status: "partially-verified"
 confidence: 0.82
 coverage: 0.54
 unknowns:
  - "long-term impact"
  - "minority cases"
 human_review_required: true
```

SECTION

21.2 不確定性不可被單一機率取代

機率通常只描述模型內部的不確定。

它未必涵蓋：

- 模型選錯；
- 適用域錯誤；
- 變數遺漏；
- 價值衝突；
- 權限問題。

  模型落差越大，越不應直接產生不可逆效果。

  定義風險：

> $$Q = R_{total} \times I_{impact} \times U_{irreversibility}$$

其中：

- R_{total}：總落差；
- I_{impact}：影響；
- $U_{irreversibility}$：不可逆程度。

  高 Q 系統應增加：

- 模擬；
- 分支；
- 人類審核；

- 小範圍試行；
- 暫停；
- 回復；
- 補償；
- 多方批准。

  若以下條件同時出現，自動化應降低或停止：

- 適用域不明；
- 落差無法估計；
- 結果高影響；
- 不可逆；
- 受影響者無法申訴；
- 目標存在價值衝突；
- 模型外案例大量出現。

  形式上：

> Automate(a)
> =
> 0

若：

> R(a)>θ_R
> ∧
> I(a)>θ_I
> ∧
> U(a)>θ_U

這不是反技術，而是承認計算代表權具有邊界。

- 模型即世界： 將資料結構視為現實本身。
- 更多資料萬能論： 認為增加資料即可消除價值與結構問題。
- 高準確率遮蔽： 用總體準確率掩蓋少數群體失效。
- 未知歸零： 把缺失、未決與不適用轉為零或預設值。
- 分類強迫： 所有對象都必須立即落入既有類型。
- 模型內真僭位： 內部一致被當成外部真實。
- 適用域漂移： 使用環境改變，模型仍被視為有效。
- 完成旗標僭位： 計算完成被當成意圖完成。
- 信心覆蓋混淆： 高信心被誤認為高世界覆蓋。
- 聚合傷害隱藏： 平均改善遮蔽個體或群體損害。
- 規範問題技術化： 把價值選擇偽裝成參數調整。
- 反事實過度自信： 將模擬方案視為真正未發生世界。
- 殘差無責任人： 限制被列出，卻沒有人負責處理。
- 免責式謙遜： 只用模糊警語轉移責任。
- 不可逆部署： 在高落差下直接產生永久效果。
- 模型外主體消失： 未被資料表示者不具申訴位置。
- 生成性未知遺忘： 未持續追蹤系統創造的新風險。
- 全面感測迷思： 為降低落差而犧牲隱私與自由。

SECTION

25.1 意圖殘差測量

比較：

- 原始對話；
- 形式規格；
- 最終模型；
- 使用者事後評價。

定義：

$$
R_I = d(I_{(stated)}, I_{(retrospective)})
$$

研究哪些類型意圖最容易在形式化中流失。

SECTION

25.2 模型覆蓋率

$$
C_M = (|relevant states represented|)/(|relevant states identified|)
$$

覆蓋率不能完全知道，但可透過事故、申訴、專家審查與模型外案例持續估計。

SECTION

25.3 未知未知事件率

追蹤部署後出現、且在設計時完全未被列入風險清單的事件。

SECTION

25.4 適用域外失效率

比較模型在原始驗證域與新域中的表現差異。

SECTION

25.5 模型外主體召回率

測量系統是否能辨識：

- 未出現在訓練資料；
- 無法使用介面；
- 不符合主要分類；
- 受到間接影響；

  的主體。

SECTION

25.6 未知保留率

檢查系統是否把未知、爭議與不適用狀態正確保存，而不是壓成預設值。

SECTION

25.7 規範衝突保留率

測量多價值衝突是否被保留給治理，或被默默壓成單一分數。

SECTION

25.8 完成落差

$$
R_{complete} = d(Complete_C, Satisfied_I)
$$

研究系統完成旗標與使用者真正目的之間的差異。

SECTION

25.9 觀測遺漏率

透過多來源資料、人工抽查、事故與申訴估計感測系統漏掉的重要狀態。

SECTION

25.10 反事實失真

比較模型預測的反事實結果與後續自然實驗、隨機試驗或替代證據。

SECTION

25.11 生成性落差追蹤

記錄系統投入使用後出現的新行為、新依賴、新制度與新風險。

SECTION

25.12 落差治理時間

測量從模型限制被發現，到：

- 被記錄；

- 被理解；
- 被修正；
- 被補救；

  所需時間。
- 表示必然包含選擇、壓縮與排除。
- 三宇宙之間不存在普遍、無損、單一、可逆的完整同構。
- 意圖落差不只是語言能力不足，也可能來自意圖尚未決定。
- 模型中的合法狀態空間小於相關現實可能空間。
- 執行結果可能偏離模型預期，即使程式本身沒有語法錯誤。
- 觀測資料只是物理世界的部分投影。
- 相同資料可形成不同解釋與後續意圖。
- 工程性落差可以降低，但結構性落差通常不能歸零。
- 規範性落差不應被更多資料或更大模型自動消除。
- 系統部署後會產生生成性落差。

  11.

> 【R_(min)>0】

對開放、多主體、持續世界系統通常成立。

- 模型內部一致不代表對外部世界完整。
- 模型適用域必須被明示、版本化與持續驗證。
- 高信心不等於高覆蓋。
- 未建模不等於不存在。
- 未回答不等於同意。

- 未分類不等於非法或無效。
- 不確定性必須成為正式狀態，而非免責備註。
- 模型落差越大、影響越高、效果越不可逆，自動化權限就應越小。

20.

> 【成熟計算不是假裝完整，】

> 【而是能精確地知道自己不完整在哪裡。】

SECTION

27.1 承接 PU-0-01

PU-0-01 建立了：

- 三個存在域；
- 三種存在判準；
- 跨宇宙同一性。

本篇進一步指出：跨宇宙同一性不能靠名稱保證，因為每次表示都會留下差異。

SECTION

27.2 承接 PU-0-02

PU-0-02 建立了：

- 四個基本映射；
- 七類耦合算子；
- 耦合殘差；
- 反身回饋。

本篇把殘差區分為：

- 工程性；
- 結構性；
- 規範性；
- 生成性。

SECTION

27.3 銜接 PU-0-04

下一篇將研究：

> 即使數位世界只是部分表示，它所建立的數位實體、身分、狀態、規則與歷史，是否仍具有真正的人工存在地位？

也就是從「模型不等於現實」推進到：

> 【不等於物理現實
> not⇒
> 沒有任何實在性】

人類不可能在沒有表示、分類與模型的情況下建立複雜程式。

沒有模型，就無法：

- 溝通；
- 推理；
- 測試；
- 預測；
- 自動化；
- 保存歷史；

- 共同建構。

  但模型一旦成功，就容易取得過度權威。

  資料庫中的欄位開始取代生活狀態。

  風險分數開始取代個人。

  完成旗標開始取代真正目的。

  模型預測開始取代世界觀測。

  這不是因為模型無用，而是因為人們忘記模型的邊界。

  因此，表示理論的核心不是反對模型，而是建立四種能力：

- 表示能力： 精確保存需要被計算的結構；
- 殘差能力： 明確保存尚未表示或不能表示的內容；
- 邊界能力： 知道模型適用與失效的條件；
- 治理能力： 在高落差、高影響與不可逆情境下保留人類及受影響者的決定權。

  本文可以收束為：

> 【Model
> =
> Represented World
> +
> Declared Boundary
> +
> Preserved Residual】

若只有第一項，模型容易僭位成世界。

若加上邊界與剩餘，它才可能成為可信任的工具。

因此，真正成熟的程式系統不應只回答：

模型知道什麼？
還必須回答：

模型不知道什麼？
哪些內容被排除了？
誰因排除而受到影響？
哪些結論只在模型內成立？
哪些世界效果尚未被證明？
何時應停止自動化並把決定交還人類？
本文的最終命題是：

> 【計算模型不是因為完整才有價值，】

> 【而是因為它能在承認不完整的前提下，】

> 【仍提供可檢驗、可修正、可反駁與可治理的表示。】

```
representation_gap_ledger:
 model_id: "eligibility-model-v3"
 version: "3.2"
 intended_purpose:
  - "identify households needing support"
  - "reduce fraudulent claims"
 represented:
  variables:
   - "reported_income"
   - "household_size"
   - "registered_address"
  time_window: "previous 12 months"
 unrepresented:
  known:
   - "informal caregiving burden"
   - "unstable housing"
   - "unreported emergency expenses"
  unknown:
   - "new crisis conditions"
 applicability:
  validated_for:
   - "households with stable documentation"
  excluded:
   - "people without fixed address"
 residuals:
  - type: "representation"
    severity: "high"
    affected_parties:
     - "undocumented applicants"
```

```
    owner: "policy-review-board"
  governance:
    human_review_required: true
    appeal_available: true
    irreversible_action_allowed: false
model_claim:
  statement: "Applicant is unlikely to qualify"
  model_confidence: 0.91
  world_coverage: 0.58
  evidence:
    observed:
      - "reported income"
      - "household registration"
    missing:
      - "current emergency costs"
      - "informal dependents"
  applicability:
    status: "borderline"
  decision:
    automation_allowed: false
    route_to: "human-review"
```

等級	特徵
R0	模型輸出被當成現實
R1	有一般免責聲明
R2	明示適用域與已知限制
R3	保存未知、爭議與不適用狀態
R4	建立落差帳本、申訴與人工覆核
R5	持續量測生成性落差並調整自動化權限
R6	表示、殘差、治理與可逆性共同編譯

- PU-0-01　三重宇宙論：意圖、計算與物理存在的基本區分
- PU-0-02　三宇宙耦合動力學：從想要、表示到世界改變
- PU-0-03　表示落差：意圖、計算模型與物理現實之間的不可完備映射
- PU-0-04　數位宇宙作為人工存在域：實體、狀態、規則與歷史

- PU-0-05　生成與限制：計算機宇宙如何創造並封閉可能性
- PU-0-06　我們到底想要什麼：計算目的、價值邊界與世界選擇

SECTION

Neo.K／EveMissLab 相關理論

- Neo.K with Aletheia，《三重宇宙論：意圖、計算與物理存在的基本區分》，2026。
- Neo.K with Aletheia，《三宇宙耦合動力學：從想要、表示到世界改變》，2026。
- Neo.K with Aletheia，《意圖中介表示：從自然語言要求到可驗證能力計畫》，2026。
- Neo.K with Aletheia，《可編譯世界：從程式執行到世界狀態演化》，2026。
- Neo.K with Aletheia，《人類可見狀態：意圖程式系統的稽核、解釋與可逆治理》，2026。
- Neo.K，《認知解構學 2.0》，2026。
- Neo.K，《語義合法性橋接》，2026。

SECTION

一般理論背景

- Korzybski, A., Science and Sanity, 1933.
- Tarski, A., “The Semantic Conception of Truth,” 1944.
- Simon, H. A., The Sciences of the Artificial, 1969.
- Box, G. E. P., “Robustness in the Strategy of Scientific Model Building,” 1979.
- Suchman, L. A., Plans and Situated Actions, 1987.
- Star, S. L. and Griesemer, J. R., “Institutional Ecology, Translations and Boundary Objects,” 1989.
- Bowker, G. C. and Star, S. L., Sorting Things Out, 1999.

- O’ Neil, C., Weapons of Math Destruction, 2016.
- Selbst, A. D. et al., “Fairness and Abstraction in Sociotechnical Systems,” 2019.
- Mitchell, M. et al., “Model Cards for Model Reporting,” 2019.
- Raji, I. D. et al., “Closing the AI Accountability Gap,” 2020.

SECTION

v0.1 — 2026-07-27

- 完成十八篇地基論文第三篇。
- 將表示定義為選擇、壓縮與排除。
- 建立意圖、表示、執行、觀測與解釋五類落差。
- 區分工程性、結構性、規範性與生成性落差。
- 提出非零最低殘差命題。
- 建立不完備表示原則與剩餘保留原則。
- 區分模型內真與世界真。
- 分析分類、尺度、反事實、規範與未知落差。
- 建立落差帳本、表示謙遜、可逆性風險與停止自動化條件。
- 加入醫療、公共福利、AI Agent、數位分身與待辦系統案例。
- 提出十八類失敗模式與十二項可證偽研究方向。
- 完成與 PU-0-04《數位宇宙作為人工存在域》的銜接。