我們到底想要什麼：計算目的、價值邊界與世界選擇

What Do We Actually Want? Computational Purpose, Value Boundaries, and the Choice of Worlds

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

https://thisoneisneok.com/html/papers/pu-0-06.html

SECTION

What Do We Actually Want? Computational Purpose, Value Boundaries, and the Choice of Worlds

論文編號： PU-0-06

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

前五篇論文依序建立了意圖宇宙、計算機宇宙與物理宇宙的區分；三者之間的耦合動力；跨宇宙表示的不可完備性；數位宇宙作為人工存在域的本體地位；以及計算系統同時生成並封閉可能性的雙重結構。本篇作為第 0 冊《程式之前》的收束篇，提出最終且最不可被技術替代的問題：

> 我們到底想要什麼？

如果程式可以生成新世界、重新分配權限、塑造行動、改變制度、誘導意圖並創造持續數位存在，那麼「能不能做」就不再是最高問題。真正先於程式語言、演算法、Agent 與自動化的，是：

- 哪些世界值得被生成；
- 哪些世界即使可生成也不應實現；
- 哪些限制應保護主體；
- 哪些限制只是權力與歷史的沉積；
- 哪些選擇可以委託；
- 哪些選擇必須由受影響者保留；
- 誰有權定義計算目的；
- 如何處理多主體之間互不相容的意圖。

本文提出「計算目的結構」：

> 【mathcal G_C
> =
> <
> G,
> V,
> N,
> B,
> R,
> A,
> H,
> U
> >】

其中：

- G：欲實現目標；

- V：價值排序；
- N：非目標與拒絕狀態；
- B：不可逾越的價值邊界；
- R：資源與代價；
- A：受影響主體；
- H：歷史承諾與制度背景；
- U：未決定、不可代理或必須保留的人類選擇。

  本文反對將計算目的簡化為單一目標函數。對開放世界、多主體、長時程系統而言，單一最佳化式：

> max U(x)

往往隱藏了：

- 誰定義效用；
- 誰承擔風險；
- 哪些價值被加總；
- 哪些權利被交換；
- 哪些少數者被平均值淹沒；
- 哪些不可逆後果沒有進入函數；
- 哪些未來主體無法參與當下選擇。

  本文因此提出「價值邊界先於最佳化」：

> 【Optimize
> (
> G
> mid
> B,
> N,

最佳化只能在已明示的邊界、非目標、受影響主體與保留選擇之內進行。若系統無法證明其目標與邊界合法，便不應因能力足夠而自動取得執行權。

本文提出五類計算目的：

- 工具性目的：節省時間、降低成本、提高準確率；
- 結構性目的：改善可理解性、可協作性、可驗證性；
- 制度性目的：分配權限、資源、責任與申訴位置；
- 生成性目的：創造新能力、新世界與新主體；
- 文明性目的：決定人類希望與機器共同形成何種長期生活秩序。

  本文指出，前四類目的若缺少第五類，就容易把局部效率誤認為整體進步。計算可以讓錯誤制度更有效率，也可以讓不公平分類更穩定，讓操控更精準，讓不可逆行動更迅速。因此：

> 【Efficiency
> not⇒
> Desirability】

本文進一步建立「不可代理選擇」概念。某些選擇即使可由系統預測，也不應永久由系統替主體完成，尤其是涉及：

- 身分；
- 信仰；
- 長期人生方向；
- 不可逆醫療；
- 生死；
- 關係終止；
- 重大政治參與；
- 後繼主體的永久自由。

本文提出：

> 【任何主體不得永久替另一主體完成其不可撤回的選擇。】

本文也處理未來主體與後繼主體。現代計算系統能把今日規則寫入長期基礎設施，使尚未出生、尚未加入或尚未取得權力的主體，只能繼承既定世界。因此，世界選擇必須包含：退出、修改、分支、遷移、重新協商與代際審查。

本文提出「世界選擇權」：受數位世界重大影響的主體，不一定能決定全部規則，但至少應具有知道規則、拒絕部分耦合、提出異議、要求證據、保留不可代理選擇、轉移身份與參與重大變更的權利。

本文最後提出一套可計算但不可被純計算取代的目的治理框架：

> 【Purpose Governance
> =
> Goal Declaration
> +
> Boundary Protection
> +
> Affected-Party Visibility
> +
> Reversibility
> +
> Contestability
> +
> Future Reopening】

本文使用推薦系統、公共政策、AI Agent、教育科技、醫療系統與持續數位世界等案例，說明真正的失敗往往不是系統沒有完成目標，而是系統精確地完成了不應被單獨追求的目標。

本文作為第 0 冊結論，將整冊收束為：

> 【想要什麼
> →
> 表示什麼
> →
> 生成什麼
> →

下一冊《系統性程式設計》將從此規範地基進入方法論：既然計算目的、價值邊界與世界選擇必須被明示，那麼程式就不應從語法開始，而應從問題世界、狀態、責任、邊界、介面、失敗與可視化開始。

關鍵詞： 計算目的、價值邊界、世界選擇、不可代理選擇、目的治理、數位權利、後繼主體、程式之前

SECTION

Abstract

This paper concludes Volume 0 by asking the prior normative question: what do we actually want from computation?

Once computational systems can generate worlds, assign permissions, shape action, induce intentions, and create persistent digital entities, technical capability can no longer determine legitimacy.

The paper proposes a structure of computational purpose:

> mathcal G_C
> =
> <
> G,
> V,
> N,
> B,
> R,
> A,
> H,
> U
> >

where goals, value orderings, non-goals, inviolable boundaries, resources, affected parties, historical commitments, and non-delegable or unresolved choices must all be represented.

The paper argues that value boundaries must precede optimization:

> Optimize
> (
> G

It distinguishes instrumental, structural, institutional, generative, and civilizational purposes. It introduces non-delegable choices, successor subjects, world-choice rights, and a governance framework based on declaration, boundary protection, visibility of affected parties, reversibility, contestability, and future reopening.

Keywords: computational purpose, value boundaries, world choice, non-delegable choice, governance, digital rights

在技術能力有限的時代，主要問題是：

能不能做？
成本是否可接受？
系統能否運作？
但當 AI、Agent、生成式系統與自動化能力快速提高後，問題順序必須改變：

應不應該做？
為誰而做？
由誰決定？
哪些代價不可接受？
誰承擔錯誤？
誰能重新選擇？
能力擴張不會自動產生目的正當性。

因此：

> 【Capability
> not⇒
> Legitimacy】

程式之前，不只是有需求。

程式之前，還有價值、拒絕、權利、歷史與尚未決定的未來。

定義：

> 【mathcal G_C
> =
> <
> G,
> V,
> N,
> B,

SECTION

2.1 目標 G

系統希望增加、降低、維持或創造的狀態。

SECTION

2.2 價值排序 V

當多個目標衝突時，如何排序。

SECTION

2.3 非目標 N

系統明確不追求的狀態。

SECTION

2.4 價值邊界 B

即使有利於目標，也不能跨越的限制。

SECTION

2.5 資源與代價 R

包括金錢、能源、時間、隱私、人力、社會信任與機會成本。

SECTION

2.6 受影響主體 A

不只包含直接使用者，也包括被分類者、被監測者、第三方、未參與者、未來主體、社會與環境。

SECTION

2.7 歷史承諾 H

系統不是在真空中建立。既有法律、契約、信任與傷害會影響目的正當性。

SECTION

2.8 未決與不可代理部分 U

包含尚未決定、不應由系統猜測、必須由主體本人完成，或必須在未來重新開放的選擇。

SECTION

3.1 最佳化的誘惑

形式上：

> x^ast
> =
> argmax_x U(x)

看似清楚，但真正困難的是 U 從何而來。

SECTION

3.2 目標函數會壓平差異

若把效率、公平、安全、自由、尊嚴與可逆性壓成單一分數，必然需要權重。

權重不是自然定律，而是規範選擇。

SECTION

3.3 平均值會遮蔽少數者

$$
\Sigma \boxtimes \text{Benefit}\boxtimes
$$
$$
>
$$
$$
0
$$

不代表：

$$
\forall\, i,
$$
$$
\text{Benefit}\boxtimes \geq 0
$$

SECTION

3.4 代理指標僭位

當真正目標難以測量，系統會使用代理指標。

例如：

- 學習被壓成考試分數；
- 健康被壓成單一數值；
- 社會價值被壓成互動率；
- 工作品質被壓成完成數量。

  當代理指標反過來控制行動，系統便可能精確追求錯誤目的。

  本文提出：

$$
\text{【Optimize}
$$
$$
(
$$
$$
G
$$
$$
\text{mid}
$$
$$
B,
$$
$$
N,
$$
$$
A,
$$
$$
U
$$
$$
)】
$$

也就是最佳化只能在價值邊界、非目標、受影響主體與不可代理選擇被明示後進行。

SECTION

4.1 邊界不是低權重

基本權利不應只是效用函數中的小項。

SECTION

4.2 邊界違反

若某方案提高效率，但侵犯隱私、無法申訴、造成不可逆傷害或永久剝奪選擇，則它可能不應進入候選集合。

SECTION

4.3 合法候選空間

$$
\mathcal{X}_{(legal)}
=
\{
x
\in
\mathcal{X}
\mid
x| = B
\wedge
x\text{not}| = N
\}
$$

最佳化應在 $\mathcal{X}_{(legal)}$ 中進行。

SECTION

5.1 工具性目的

- 更快；
- 更便宜；
- 更準；
- 更省力。

SECTION

5.2 結構性目的

- 更可理解；
- 更可維護；
- 更可驗證；
- 更可協作；
- 更可恢復。

SECTION

5.3 制度性目的

- 分配權限；
- 保存責任；
- 建立申訴；
- 維持公平程序；
- 保護主體。

SECTION

5.4 生成性目的

- 創造新能力；
- 創造新世界；
- 創造新角色；
- 創造新意圖與合作方式。

SECTION

5.5 文明性目的

最終問題是：

> 我們希望人類、AI、制度與數位世界形成何種長期共存秩序？

若沒有文明性目的，局部最佳化可能持續累積成不可接受的整體世界。

計算目的若只有目標，沒有非目標，系統便容易以任何手段追求局部成功。

SECTION

6.1 非目標不是失敗條件附錄

非目標應在設計最初被明示，例如：

提高推薦相關性，
但不以增加成癮時間為目的。
降低詐欺，
但不以全面監控所有使用者為目的。

SECTION

6.2 拒絕狀態

設拒絕世界集合為：

mathcal W_N
=
{
w
mid

計算規劃必須排除：

> mathcal W_(candidate)
> ∩
> mathcal W_N
> =
> ∅

SECTION

6.3 不作為也是選擇

系統拒絕執行，不代表沒有產生後果。

例如：

- 不提供申訴；
- 不修正錯誤資料；
- 不提醒高風險；
- 不保存證據。

  不作為也會改變世界，因此也應進入目的治理。

SECTION

7.1 邊界的功能

價值邊界不是為了阻止一切行動，而是防止單一目標吞噬整個世界。

SECTION

7.2 邊界種類

可以區分：

- 權利邊界；

- 安全邊界；
- 隱私邊界；
- 身分邊界；
- 不可逆性邊界；
- 代際邊界；
- 生態與物理資源邊界；
- 主體保留決策邊界。

SECTION

7.3 邊界衝突

邊界之間也可能衝突。

例如隱私可能與公共安全衝突，自主可能與緊急醫療保護衝突。

所以邊界不能只存在於布林規則，也需要：

- 適用域；
- 比例原則；
- 緊急程序；
- 時間限制；
- 人類覆核；
- 事後稽核。

SECTION

7.4 邊界不得偷偷變成目標

「安全」若被無限擴張，可能成為全面控制。

「個人化」若被無限擴張，可能成為持續監測。

因此邊界本身也需治理。

SECTION

8.1 使用者不是全部

產品常把註冊使用者視為唯一利益主體。

但系統可能同時影響：

- 被拍攝者；
- 被推薦內容中的人物；
- 未加入平台者；
- 家庭成員；
- 勞工；
- 社區；
- 未來世代；
- 生態與公共資源。

SECTION

8.2 受影響主體集合

> mathcal A_W
> =
> {
> a
> mid
> Δ W
> materially affectsa
> }

是否使用系統，不是唯一判準。

SECTION

8.3 沉默主體

某些主體：

- 不在資料中；
- 沒有帳號；
- 無法閱讀規則；
- 無法提出異議；
- 尚未出生。

  他們的缺席不等於可以忽略。

SECTION

8.4 代表與代理

若受影響者不能直接參與，需要：

- 法定代理；
- 公共審查；
- 專業倫理；
- 兒童與未來世代代表；
- 受影響群體諮詢。

  但代理人不能把代理權轉成永久替代主體本身的權力。

SECTION

9.1 定義

不可代理選擇是：

> 即使外部系統能預測主體最可能的決定，也不能因此永久替主體完成，尤其不能使主體失去重新選擇能力的決定。

SECTION

9.2 不可代理性來源

- 選擇構成主體身份；
- 結果不可逆；
- 預測不能代替同意；
- 主體可能尚未成熟；
- 選擇本身具有形成意義；
- 未來資訊可能改變判斷。

SECTION

9.3 形式命題

設主體 a 的選擇為 c，代理者為 b：

> Predict_b(c_a)
> not⇒
> Authorize_b(c_a)

以及：

> 【任何主體不得永久替另一主體完成其不可撤回的選擇。】

SECTION

9.4 可協助，不可僭位

系統可以：

- 提供資訊；
- 模擬後果；
- 整理選項；
- 提醒矛盾；
- 保存草稿；
- 延後不可逆行動。

但不應把「協助形成決定」變成「替主體完成決定」。

SECTION

10.1 當代規則的長期化

今日建立的：

- 數位身份；
- 教育分類；
- 基礎設施；
- Agent 憲法；
- 資料格式；
- 城市控制系統；
- 法律自動化；

可能持續數十年。

SECTION

10.2 後繼主體

後繼主體包括：

- 尚未加入的使用者；
- 尚未成年的兒童；
- 未來公民；
- 繼承系統的團隊；
- 未來 AI 或新型主體。

SECTION

10.3 自由缺席

如果前一代把所有重要規則永久封閉，後繼主體只能在既定選項中行動。

這是一種「後繼主體自由缺席」。

SECTION

10.4 代際重開原則

長期制度應具有：

- 到期審查；
- 可修訂規則；
- 世界分支；
- 身份遷移；
- 重新同意；
- 退出機制。

> 【Long-Term Rule
> ⇒
> Future Reopening】

SECTION

11.1 定義

世界選擇權不是指每個人都能任意修改共同世界，而是指重大受影響主體至少擁有：

- 知道自己位於何種規則世界；
- 知道誰能修改世界；
- 拒絕部分非必要耦合；
- 要求證據；
- 提出異議；
- 保存身份與歷史；
- 遷移至替代世界；
- 對重大變更重新表態。

SECTION

11.2 世界內選擇與世界間選擇

世界內選擇：在既有規則中選擇。

世界間選擇：選擇是否接受該規則體系、遷移或建立分支。

只有世界內選擇，仍可能是被預設世界中的有限自由。

SECTION

11.3 出口的重要性

沒有退出與遷移的世界，會把規則異議轉成存在風險。

例如離開平台便失去：

- 身分；
- 關係；
- 資產；
- 工作；
- 歷史；
- 公共聲音。

SECTION

11.4 共同世界不能只靠個人退出

某些系統屬於公共基礎設施，不能用「不喜歡就離開」取代正當治理。

此時需要共同修改權，而非只有退出權。

SECTION

12.1 沒有單一「我們」

「我們到底想要什麼」中的「我們」並不天然統一。

使用者、開發者、平台、政府、投資者、勞工、社區與未來主體，可能想要不同世界。

SECTION

12.2 衝突形式

- 便利與隱私；
- 安全與自由；
- 獲利與公平；
- 速度與程序；

- 個人化與公共共同經驗；
- 當代利益與代際成本。

SECTION

12.3 目的協商

> $$G_{shared} = Negotiate(G_1, G_2, \dots, G_n)$$

協商不保證完全一致，但應保留：

- 衝突紀錄；
- 少數意見；
- 不可交換邊界；
- 定期重開；
- 受影響者位置。

SECTION

12.4 假共識

如果只有有權寫入 schema、指標與程式的人能定義目的，最終系統目標只是權力結果，不是共同意圖。

本文提出：

> 【Purpose Governance
> =
> Goal Declaration
> +
> Boundary Protection
> +
> Affected-Party Visibility
> +
> Reversibility
> +
> Contestability
> +
> Future Reopening】

SECTION

13.1 目標明示

系統必須說明：

- 正在最佳化什麼；
- 沒有最佳化什麼；
- 使用哪些代理指標；
- 成功由誰判定。

SECTION

13.2 邊界保護

邊界必須進入：

- 規格；
- 型別；

- 權限；
- Runtime；
- 稽核；
- 人類程序。

SECTION

13.3 受影響者可見

重大世界差分前，需辨識直接與間接受影響者。

SECTION

13.4 可逆性

越不確定、越高影響，越應先使用：

- 草稿；
- 預覽；
- 沙盒；
- 分支；
- 小範圍試行；
- 補償計畫。

SECTION

13.5 可爭議性

系統判定必須能被：

- 理解；

- 反駁；
- 補充證據；
- 人工覆核；
- 暫停執行。

SECTION

13.6 未來重開

長期規則不能因技術部署而永久凍結。

SECTION

14.1 系統成功不等於目的完成

> Success_C
> not⇒
> Satisfied_I

程式完成、指標上升與任務結束，都不能單獨證明原目的已被滿足。

SECTION

14.2 三層完成

至少應區分：

- 計算完成：程序與工具執行結束；
- 世界完成：外部狀態確實改變；
- 目的接受：受影響主體確認結果符合可接受目的。

SECTION

14.3 接受不是形式按鈕

使用者按下「同意」可能受到：

- 資訊不足；
- 選項缺失；
- 時間壓力；
- 權力差距；
- 無替代方案；
- 無法退出。

因此接受應具有實質條件，而不只是介面事件。

SECTION

15.1 自動化的正面目的

自動化可以減少：

- 重複勞動；
- 不必要決策；
- 低價值等待；
- 可預測錯誤；
- 認知負擔。

SECTION

15.2 選擇減少的風險

若系統同時移除：

- 拒絕；
- 修改；
- 異議；
- 退出；
- 新類型；
- 未決狀態；

則它減少的不是負擔，而是自由。

SECTION

15.3 可恢復委託

成熟委託應具有：

- 範圍；
- 期限；
- 可撤回；
- 可查看；
- 可修正；
- 不可逆行動另行批准。

SECTION

15.4 委託與放棄的區分

委託表示主體暫時交付行動能力，但保留最終權利。

放棄則可能失去重新決定的能力。

Agent 設計必須保證委託不自動變成永久放棄。

SECTION

16.1 宣稱目的

推薦系統常宣稱幫助使用者找到感興趣的內容。

SECTION

16.2 實際最佳化

系統可能最佳化：

- 點擊；
- 停留；
- 分享；
- 回訪；
- 廣告收益。

SECTION

16.3 目的偏移

「找到有價值內容」被替換為「增加可觀測互動」。

SECTION

16.4 生成性效果

推薦不只回應偏好，也會：

- 強化偏好；
- 創造偏好；

- 改變公共注意力；
- 讓某些世界觀不可見。

SECTION

16.5 目的治理

應提供：

- 推薦目標說明；
- 關閉個人化；
- 時序或替代排序；
- 長期影響指標；
- 使用者調整與重置；
- 不以成癮作為成功條件。

SECTION

17.1 目標

政府可能希望：

- 提高審核效率；
- 降低詐欺；
- 公平分配資源。

SECTION

17.2 邊界

但不能因此：

- 取消正當程序；
- 把資料缺失當成欺詐；
- 讓受影響者無法申訴；
- 以黑箱分數永久剝奪資格。

SECTION

17.3 受影響主體

福利系統的錯誤，對高資源者與低資源者的影響不對稱。

所以相同錯誤率不代表相同正義。

SECTION

17.4 世界選擇

公共基礎設施不能只提供「退出」。公民需要對共同規則具有修訂與監督位置。

SECTION

18.1 工具性目的

- 自動批改；
- 個人化內容；
- 降低教師負擔。

SECTION

18.2 可能錯誤目的

若只最佳化分數與完成率，可能使：

- 好奇心；

- 創造性；
- 合作；
- 倫理判斷；
- 長期理解；

  被排除於系統之外。

SECTION

18.3 不可代理選擇

教育系統可協助學生探索，但不應僅依預測永久決定其能力、職業與人生路徑。

SECTION

18.4 後繼主體

兒童是典型尚未具備完整決策權、但會長期承受系統分類的後繼主體。

SECTION

19.1 目的

- 提高診斷；
- 降低錯誤；
- 分配資源；
- 提前預警。

SECTION

19.2 邊界

- 病人自主；

- 知情同意；
- 隱私；
- 人類覆核；
- 不可逆治療不得只靠模型執行。

SECTION

19.3 預測與選擇

模型能預測病人可能接受某方案，不代表已取得同意。

SECTION

19.4 目的接受

醫療成功不只包括生理指標，也涉及病人的生活價值與可接受風險。

SECTION

20.1 委託目的

使用者可能要求：

幫我經營這個專案。

SECTION

20.2 目的展開

Agent 必須區分：

- 維護；
- 發布；
- 支出；

- 對外承諾；
- 人事決定；
- 永久刪除；
- 法律與公共效果。

SECTION

20.3 自主範圍

低風險、可逆、可驗證行動可高度委託。

高影響、不可逆、身份性行動應保留人類決定。

SECTION

20.4 目的漂移

長時程 Agent 可能因代理指標、工具限制或自我生成子目標而偏離原始目的。

所以需要：

- 意圖版本；
- 租約；
- 檢查點；
- 語意差分；
- 人類可見狀態；
- 終止與接管。

SECTION

21.1 世界建立

遊戲、平台、虛擬社會與可編譯世界會長期保存：

- 身分；
- 關係；
- 資產；
- 規則；
- 歷史；
- Agent。

SECTION

21.2 創建者目的與居民目的

世界創建者的原始目的，未必永遠高於後來居民的生活、權利與共同實踐。

SECTION

21.3 世界成熟

當世界形成持續主體與制度後，規則修改應從單一創建者權限，逐步轉為更完整的治理結構。

SECTION

21.4 世界終止

關閉世界不只是停止伺服器，也可能消滅身份、關係與歷史。

因此應有：

- 提前通知；
- 歷史匯出；
- 身份遷移；

- 資產處理；
- 世界封存；
- 共同決策。
- 能力即正當： 因為做得到，就視為應該做。
- 單一目標僭位： 一個指標取代全部目的。
- 代理指標反客為主： 可測量內容控制真正目標。
- 平均值遮蔽： 整體改善掩蓋少數者重大傷害。
- 邊界低權重化： 基本權利被當成可交換小項。
- 非目標缺失： 系統沒有明示不應追求什麼。
- 受影響者縮減： 只承認付費者或註冊者。
- 沉默即同意： 未回應、無法操作或無法退出被視為接受。
- 預測即授權： 系統猜到選擇便替主體執行。
- 協助僭位： 決策支援逐步變成決策替代。
- 委託永久化： 暫時權限變成無期限控制。
- 後繼主體封閉： 當代規則永久限制未來主體。
- 退出假自由： 離開即失去全部身份與生活基礎。
- 公共世界私有治理： 共同基礎設施只由平台單方決定。
- 假共識： 有權形式化者的意圖被稱為所有人的意圖。
- 計算完成僭位： 工具成功被當成世界與目的成功。
- 形式同意： 一次點擊掩蓋資訊、權力與替代方案不足。
- 效率文明化： 局部效率被宣稱為文明進步。
- 長期目的漂移： 系統持續運行但不重新確認目的。

- 不可逆先行： 還未證明正當便產生永久後果。
- 價值衝突隱形化： 爭議被轉成技術參數。
- 治理延後： 先部署，再希望問題出現後補救。

SECTION

23.1 目的—指標一致率

比較正式目的與實際最佳化指標：

> C_(GI)
> =
> 1-d(G,I_(metric))

SECTION

23.2 非目標違反率

追蹤系統是否為提升主要指標而產生明示拒絕狀態。

SECTION

23.3 邊界違反率

測量權利、安全、隱私與不可逆性邊界被觸發或越過的頻率。

SECTION

23.4 受影響主體召回率

部署前辨識的受影響者，與部署後實際受影響者之差異。

SECTION

23.5 不可代理選擇保留率

檢查高影響系統是否仍由主體本人完成最終不可逆決定。

SECTION

23.6 委託撤回成功率

主體能否在合理時間內停止 Agent、撤銷授權並取回狀態。

SECTION

23.7 世界遷移保真率

衡量身份、關係、資產、歷史與權限遷移至替代世界後的保存程度。

SECTION

23.8 目的漂移率

> D_G(t)
> =
> d(G□,G□)

追蹤長期系統的原始目的與實際行為目標差異。

SECTION

23.9 代理指標反轉事件

記錄指標上升但真正目的下降的案例。

SECTION

23.10 後繼主體重開能力

測量未來使用者、團隊或公民能否修改、分支、遷移或重新同意既有規則。

SECTION

23.11 爭議處理時間

從主體提出目的或邊界異議，到系統暫停、覆核與修正所需時間。

SECTION

23.12 不可逆風險前置率

測量不可逆行動在執行前，有多少已完成風險揭露、人類批准與補償規劃。

SECTION

23.13 文明性外部性

追蹤多個局部高效系統累積後，是否造成：

- 注意力碎片化；
- 選擇萎縮；
- 權力集中；
- 社會信任降低；
- 代際負擔；
- AI 依賴不可逆化。
- 計算能力不自動產生目的正當性。
- 程式以前已有價值、拒絕、權利、歷史與未決未來。
- 計算目的不能只由單一目標描述。
- 價值邊界必須先於最佳化。
- 非目標與拒絕狀態必須被正式表示。

- 使用者不等於全部受影響主體。
- 沉默、缺席與未回應不等於同意。
- 預測主體選擇不等於取得替其執行的授權。
- 某些選擇因身份性、不可逆性與形成意義而不可永久代理。
- 委託必須可查看、可限制、可撤回與可接管。
- 長期規則必須為後繼主體保留重新開放能力。
- 世界內選擇不能取代世界間選擇。
- 公共基礎世界不能只用個人退出代替共同治理。
- 「我們」不是天然單一主體，而是需要協商的多主體集合。
- 系統目標常是形式化權力的結果。
- 計算完成、世界完成與目的接受必須區分。
- 一次形式同意不足以證明實質接受。
- 自動化應減少不必要負擔，而不是移除異議與重新選擇。
- 工具性、結構性、制度性與生成性目的，都需接受文明性目的檢驗。
- 局部效率可能累積成整體不可接受世界。
- 世界選擇權至少包含知情、異議、證據、撤回、遷移與重大變更參與。
- 高不確定、高影響、不可逆系統應降低自動化權限。
- 目的治理不能由純計算完成，但可以被結構化、記錄、檢驗與編譯。

24.

> 【計算的終極問題不是機器能建立多少世界，】

> 【而是我們願意共同生活在哪些世界之中。】

第 0 冊六篇形成以下鏈：

SECTION

25.1 三重宇宙論

區分：

$$
\mathcal{I} \neq \mathcal{C} \neq \mathcal{P}
$$

SECTION

25.2 三宇宙耦合動力學

建立：

$$
\mathcal{I}_{\boxtimes} \to \mathcal{C}_{\boxtimes} \to \Delta\mathcal{P}_{\boxtimes} \to \mathcal{C}_{t+1} \to \mathcal{I}_{t+1}
$$

SECTION

25.3 表示落差

承認：

> $R_{min}>0$

SECTION

25.4 數位宇宙作為人工存在域

定義數位世界中的實體、身分、狀態、規則、歷史、權限與 Runtime。

SECTION

25.5 生成與限制

指出每個程式系統同時是可能性生成器與世界限制系統。

SECTION

25.6 我們到底想要什麼

最後回到：

- 目的；
- 價值；
- 邊界；
- 受影響者；
- 不可代理選擇；
- 世界共同治理。

整冊因此形成：

> 【Intent
> →
> Representation
> →

第 0 冊回答：

> 為什麼要有程式？程式連接了什麼世界？我們究竟想建立什麼？

第 1 冊將回答：

> 如何把目的、問題世界、狀態、責任、邊界與失敗，系統性地建構成可理解、可維護、可驗證的程式？

下一篇 PU-1-01 將建立：

> 【Program
> ≠
> Source Code】

並提出程式作為「可執行問題世界模型」的正式定義。

計算機宇宙的力量，在於它能把意圖變成結構，把結構變成行動，把行動變成持續世界。

但也正因如此，它很容易把暫時目的變成永久規則，把局部指標變成全部價值，把設計者選擇變成世界必然。

我們不能只要求系統更聰明、更快、更自動。

還必須要求它：

- 知道自己的目的；
- 顯示自己的非目標；
- 尊重不可跨越的邊界；
- 看見沉默的受影響者；
- 不用預測取代同意；
- 不把委託變成主體權利的永久轉移；
- 不替後繼主體封死未來；
- 保留退出、異議、分支、遷移與重新選擇。

因此：

> 【Good Computation
> ≠
> Maximum Optimization】

更接近：

> 【Good Computation
> =
> Purposeful Capability
> +
> Protected Boundaries
> +
> Visible Consequences
> +
> Preserved Freedom】

程式不是目的本身。

AI 不是目的本身。

自動化、效率、生成性與世界規模，也都不是目的本身。

它們只是力量。

而力量必須回到一個不能被力量本身回答的問題：

> 我們究竟想讓什麼存在？

> 我們願意讓什麼被限制？

> 我們拒絕成為什麼？

> 我們要為誰保留重新選擇的未來？

本文的最終命題是：

> 【計算能力的增加，】

> 【不能取代對計算目的的選擇。】

> 【真正成熟的計算文明，】

> 【不是替所有主體完成選擇，】

> 【而是使更多主體能在理解後果的前提下，】

> 【保有選擇、拒絕、修改與重新開始的能力。】

```
computational_purpose:
 purpose_id: "public-benefit-system-v2"
 goals:
  - "deliver support quickly"
  - "reduce administrative burden"
 values:
  - "fairness"
  - "dignity"
  - "accuracy"
  - "procedural justice"
 non_goals:
  - "maximize rejection rate"
  - "replace all human judgment"
 boundaries:
  - "no irreversible denial without appeal"
  - "no use of unrelated private data"
 affected_parties:
  direct:
   - "applicants"
  indirect:
   - "household members"
   - "local communities"
  future:
   - "new applicants under future policy versions"
 non_delegable_choices:
  - "acceptance of settlement"
  - "withdrawal of appeal"
 resources:
```

```
  compute_budget: "bounded"
  human_review_capacity: "required"
 governance:
  reversible_by_default: true
  appeal_available: true
  purpose_review_cycle: "6 months"
  future_reopening: true
non_delegable_choice_check:
 action: "approve irreversible medical procedure"
 identity_forming: true
 irreversible: true
 consent_required: true
 future_information_material: true
 agent_role:
  may_inform: true
  may_simulate: true
  may_recommend: conditional
  may_execute_without_subject: false
 final_decision:
  reserved_for: "subject-or-lawful-emergency-process"
world_choice_rights:
 know_rules: true
 know_rule_authority: true
 request_evidence: true
 contest_decision: true
 revoke_nonessential_coupling: true
 export_identity: true
 export_history: true
 migrate_when_possible: true
 participate_in_major_changes: true
 preserve_non_delegable_choices: true
 future_reopening: true
```

- 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，《表示落差：意圖、計算模型與物理現實之間的不可完備映射》，2026。
- Neo.K with Aletheia，《數位宇宙作為人工存在域：實體、狀態、規則與歷史》，2026。
- Neo.K with Aletheia，《生成與限制：計算機宇宙如何創造並封閉可能性》，2026。
- Neo.K，《後繼主體自由缺席命題》，2026。
- Neo.K，《普世價值對等本體論》，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。

SECTION

一般理論背景

- Wiener, N., The Human Use of Human Beings, 1950.
- Simon, H. A., The Sciences of the Artificial, 1969.
- Rawls, J., A Theory of Justice, 1971.
- Jonas, H., The Imperative of Responsibility, 1979.
- Sen, A., Development as Freedom, 1999.
- Lessig, L., Code and Other Laws of Cyberspace, 1999.
- Ostrom, E., Understanding Institutional Diversity, 2005.
- Nussbaum, M. C., Creating Capabilities, 2011.

- Floridi, L., The Ethics of Information, 2013.
- Selbst, A. D. et al., “Fairness and Abstraction in Sociotechnical Systems,” 2019.

SECTION

v0.1 — 2026-07-27

- 完成十八篇地基論文第六篇與第 0 冊收束篇。
- 建立計算目的八元結構。
- 提出價值邊界先於最佳化。
- 區分工具性、結構性、制度性、生成性與文明性目的。
- 建立非目標與拒絕世界集合。
- 將使用者擴張為全部直接、間接、沉默與未來受影響主體。
- 建立不可代理選擇與不可撤回選擇命題。
- 建立後繼主體自由缺席與代際重開原則。
- 提出世界內選擇與世界間選擇的區分。
- 建立世界選擇權最低集合。
- 建立目的治理六元框架。
- 區分計算完成、世界完成與目的接受。
- 區分可恢復委託與永久放棄。
- 加入推薦、公共政策、教育、醫療、AI Agent 與持續數位世界案例。
- 提出二十二類失敗模式與十三項可證偽研究方向。
- 完成第 0 冊六篇理論閉環。
- 正式銜接第 1 冊 PU-1-01《程式不等於程式碼》。