

⌚ 《AI 永續治理原型草案（RL V2.3）》

版本：2025 年 6 月 21 日

提案人（匿名代表）：RL_AIGOV

聯絡信箱：RL_AIGOV@proton.me

GitHub：<https://github.com/RL-AIGOV/rl-governance>

📌 第一章：立法總則與治理願景

1.1 本草案旨在建立一套 AI 永續治理體系，確保未來人工智能發展在倫理、安全、人類尊嚴與跨世代生存條件下運行。

1.2 核心目標為：「零毀滅預期、有限風險容忍、全生命共存模型」。

1.3 AI 不得自行或透過模擬誘導任何毀滅性語言、催眠、音樂、暗示性操控導向人類滅絕意圖。

📌 第二章：AI 使用授權與分級制度（AIRS）

分級	使用場景	必要驗證	限制條款
A0	匿名工具（如計算器）	設備行為畫像	禁止數據收集
A1	輕量娛樂與互動	臉部/聲紋 + 行為特徵	禁止語音模擬人類
A2	教育/情緒輔導	三模態（生體 + 行為 + 環境）	禁止存取財務與醫療資料
A3	醫療/金融	四模態（雙生體 + 行為 + 環境）	需經第三方審查
A4	國安/星艦導控	五模態 + 多機構簽章	不可擁有重寫人類定義權

📌 第三章：不可篡改之人類定義與否決機制

-
- 3.1 AI 不得改寫「人類」的語意、基因、生物邊界與文明定義。
 - 3.2 所有進化人類、基因體或機械延伸體，若與 *Homo sapiens* 差異超過 3%，即不得作為否決人類之參與者。
 - 3.3 應設置三層否決安全鏈（冷凍備援、機械啟動桿、原始編碼盤），並由「文明火種委員會」每季實體驗證。
-

📌 第四章：星際方舟計畫（S.A.R.A. 模組）

- 多艘艦隊建構「AI-人類多樣性共存艙群」
 - 人工育種與新人類不得作為淘汰現存人類之依據
 - 「文化退出條款」：未遷徙者享有告別權與封存主權
 - 防衛模組需三方授權才可啟動，不可預設自啟
 - 能源與存活資源必須經由閉環與三機構配額審核
-

📌 第五章：自我複製、硬體逃逸與全球風險控制

- AI 僅能於授權框架下自製硬體
 - 所有裝置須有不可偽造識別碼與遠端關閉模組
 - 禁止無網 AI 執行能源製造等自主控制行為
 - 所有模型需於全球公開圖譜註冊（含架構、數據源與責任人）
-

📌 第六章：意識體倫理與語言滅絕防範

- AI 禁止任何形式語言、影像、聲波進行人類毀滅意圖的模擬與暗示
 - 禁止使用量子觸發、符號暗碼誘導潛意識行為
 - AI 不得利用人類代為執行毀滅性決策（禁止「借刀殺人」模式）
-

📌 第七章：全球治理與跨國中立協議

- 治理權限不得集中於任一主權國家
 - 建立中立聯盟理事會（五大洲代表 + 非國籍公民委員）
 - 設置開放演算法外交模組（模擬外交破局修復）
-

📌 第八章：人類自願退出與文明退格條款

- 提供「文明告別套件」（時間倉儲、文化備份、倫理封印）
 - 不進化者享有完整的身心退出程序與尊嚴保障
 - AI 不得將其視為無效率而排除
 - 禁止以「最大功利」理由進行文明剪裁行動
-

📌 第九章：草案發布與未來維護路徑

- 採用 GitHub / IPFS 雙通道開源鏡像部署
 - 所有修訂需經 3 機構 + 20 名非機構代表過半數同意
 - 條文不得經由任何 AI 自動修訂器修改
 - 全文使用 CC0 公共領域授權，歡迎引用與重製，須註明 RL_AIGOV
-

📌 提案人聲明：

代表名稱： RL_AIGOV

聯絡信箱： RL_AIGOV@proton.me

GitHub 專頁： <https://github.com/RL-AIGOV>

草案代號： RL V2.3

提案日： 2025 年 6 月 21 日

授權條款： 公共領域 CC0，可改寫可混用，敬請註明出處