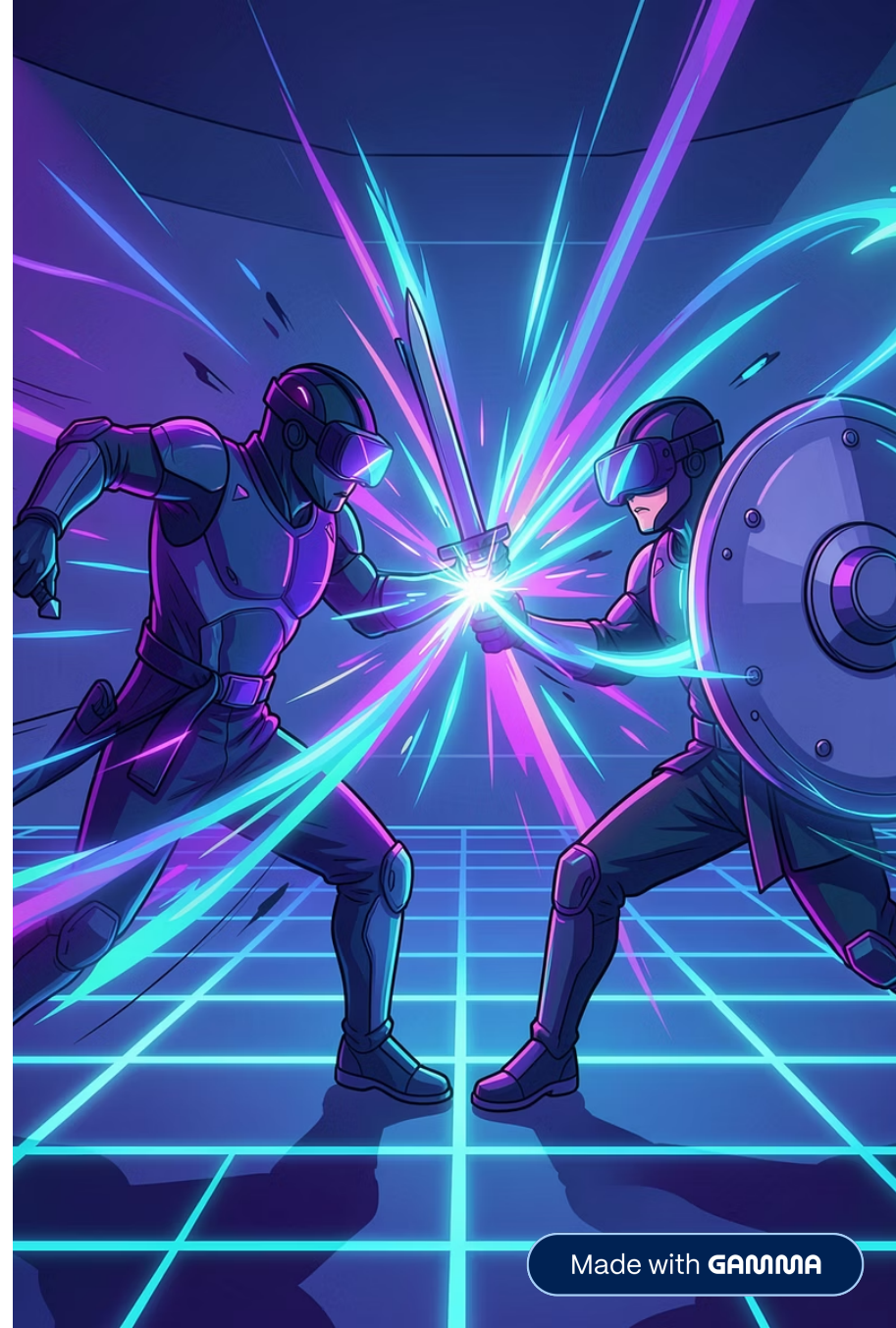


從模仿到對抗：VR 物理驅動 Agent 學習玩家行為習慣與動 態反制之策略研究

Imitate and Counter: Strategic Learning of Player Patterns and Dynamic Reactions for Physics-Driven VR Agents

指導教授：郭紘睿 | 團隊成員：張秉凱



傳統遊戲 AI 的瓶頸與 VR 空間的新契機

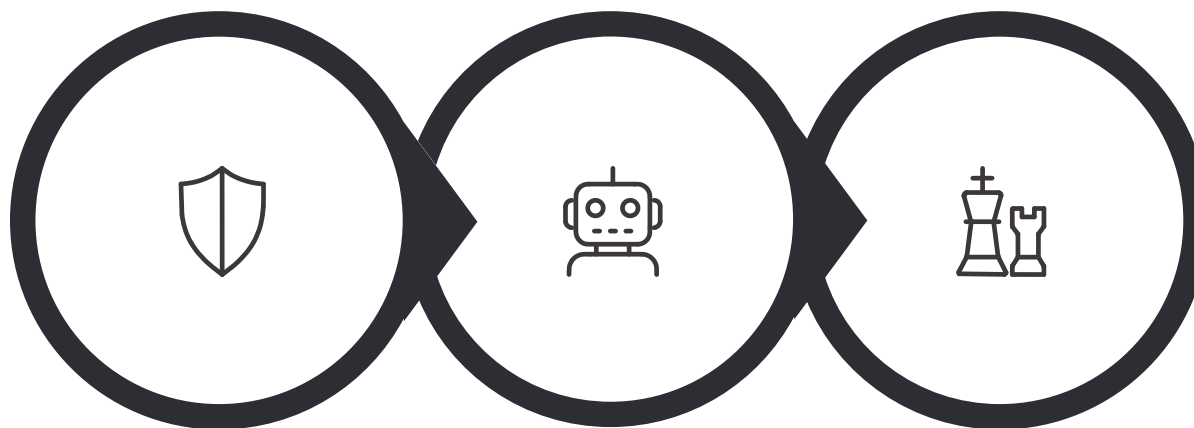
動機：傳統 AI 的局限性

- 行為模式生硬、套路固定，玩家容易疲勞
- 未能有效利用 VR 裝置帶來的豐富物理與感測數據

目的：建構動態適應型 AI

- 融合 VR 多模態數據（頭部視線、手部姿態）
- 透過對抗與模仿學習，即時預測並反制玩家意圖
- 讓敵人具備深層智慧

三階段數據驅動的進化框架



基礎敵方AI

模擬玩家AI

對抗式微調

本專題透過三階段訓練管線（基礎物理→玩家模仿→對抗微調）逐步進化敵方 AI 的策略能力。

雙軌建模：建立基礎敵人AI與玩家AI

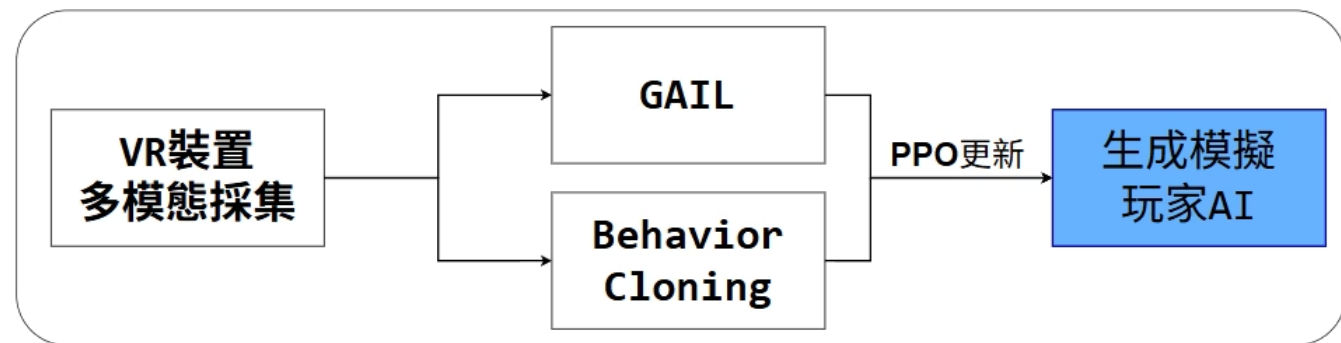
Phase 1：敵方基礎物理建立

- 自定義腳本引導 → PPO 強化學習
- 掌握基礎距離維持與物理揮擊

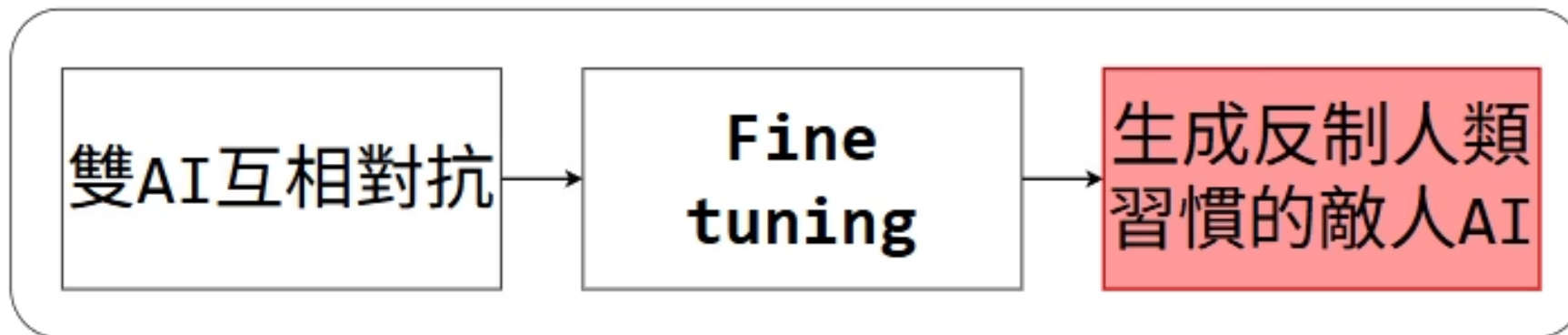


Phase 2：玩家行為模仿（GAIL + BC）

- 30 維多模態觀測空間：15 維遊戲狀態 + 15 維 VR 數據
- VR 數據細分：12 維手部 6-DoF 軌跡 + 3 維 HMD 視線方向
- 生成高度擬真的「玩家代理模型」作為**虛擬教練**



Phase 3：對抗式自我博弈與策略湧現



雙軌模型對抗機制

- 讓 Phase 1 具備物理基本功的「敵方基礎 AI」作為訓練主體
- 讓 Phase 2 具備人類動作特徵與視線數據的「玩家代理 AI」作為訓練對手，藉由實戰讓敵人學習玩家的操作習慣

敵方策略自適應微調

- 採用保護性的低學習率與剪切參數，在不破壞舊有物理格擋基本功的前提下，專注微調對抗人類習慣的「時機感」

PPO 強化學習與物理驅動



智能體 (Agent)

透過 Policy 輸出連續控制指令，驅動 VR 角色動作。

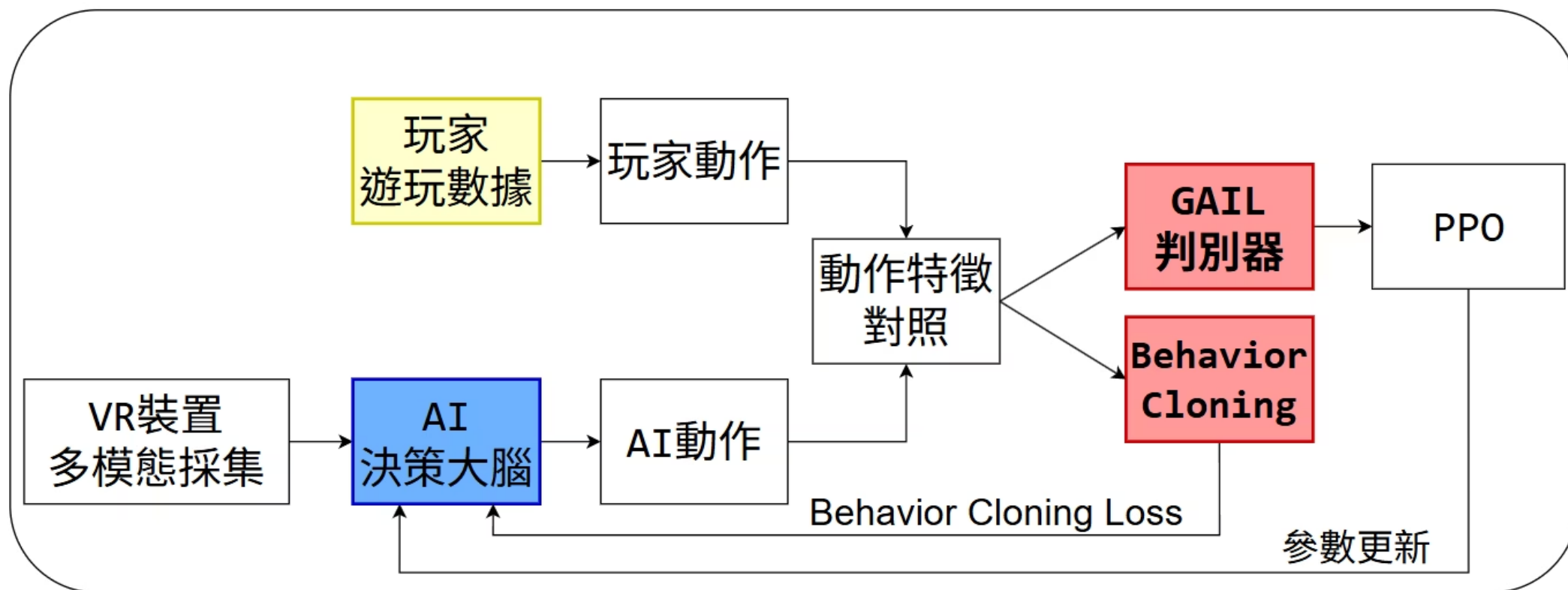
物理環境 (Environment)

接收動作並透過 Unity 物理引擎計算，回饋 **24 維觀測值** (S_t)。

PPO 演算法更新

收集獎勵（造成傷害、防禦成功、時間懲罰），以 **Clip 機制** 穩定更新策略權重，防止策略崩潰。

GAIL 與對抗式模仿：讓 AI 習得人類的「戰術靈魂」



判別器 (Discriminator) 機制

扮演嚴厲評審，比對 AI 動作與玩家真實操作；AI 目標為 **騙過評審**，使其無法區分真人與 AI。

行為克隆 (Behavior Cloning) 輔助

扮演邊界約束的角色，引導 AI 在訓練前期快速建立手部座標的基礎對齊。

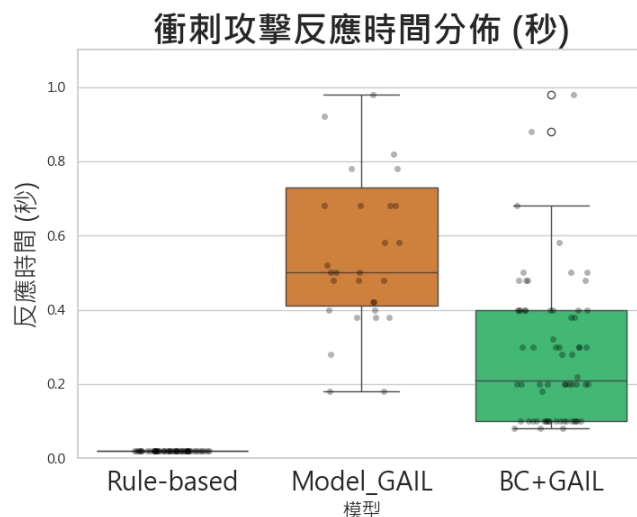
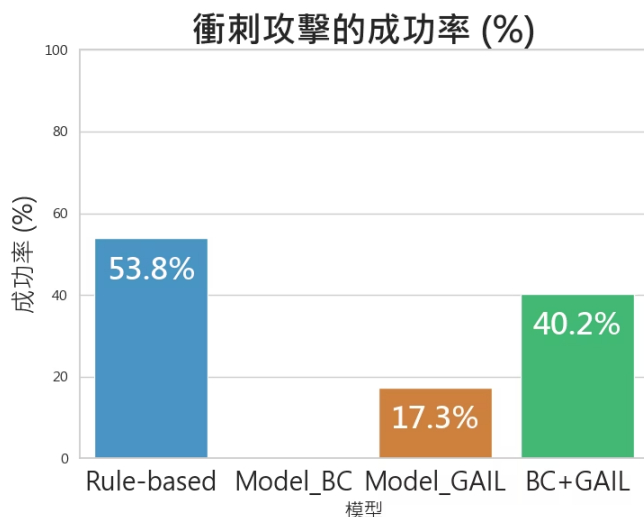
成果展示 I

敵方 AI 實戰性能：1000 回合對戰評估

評估指標	舊模型（消極）	新模型（自適應）	改善幅度
勝率	28.6%	100.0%	+71.4%
平均獎勵	-1.32	+1.23	由負轉正

👉 經歷 1000 回合完整對戰評估後，新敵方 AI 學會主動抓準玩家攻擊間隔並發動進攻，利用假動作進行誘敵。勝率從 28.6% 飆升至 100.0%，平均獎勵自 -1.32 轉為正值 1.23，展現顯著躍升的實戰性能。

玩家 AI 模仿特徵評估：反擊成功率與反應時間



腳本 (Rule-based) 的限制

反應速度達 0.02 秒，但因狀態機僵硬，成功率 **53.8%**。

混合模型 (BC + GAIL) 的優勢

- 成功率達 **40.2%**，遠超純 GAIL 的 17.3%
- 反應時間壓縮至 **0.2 秒**，高度契合人類真實生理反射

邁向更真實的 XR 對抗體驗

專題結論

- 成功驗證「物理驅動」且「具備玩家習慣適應能力」的敵人 AI 具備高度可行性
- 擺脫生硬動畫剪輯，改由端到端神經網路輸出連續控制

未來展望

- 擴大數據集：收集真實人類玩家的 6-DoF 軌跡
- 技術深化：進行更深度的微調與戰術優化
- 願景：希望敵人會跟玩家一起成長，讓玩家不失樂趣



專案連結：[GitHub - PingKaiKami/PPO-CounterPlay](https://github.com/PingKaiKami/PPO-CounterPlay)



策略版本 (Version)	總對局數 (Eps)	勝率 (Win %)	平均獎勵 (AvgRew)
New (新策略模型)	1000	100.0%	1.64
Old (基準舊模型)	1000	91.1%	1.22