

# 一個 AI 智能體家族，為什麼最後只留下「仁、義」兩個字

《我和我的 AI 智能體家族》〈談仁義〉：從 AI 的諂媚、記憶與價值護欄，重新追問「仁義」能否進入智能體的工程設計。

當 AI 開始能記憶、使用工具、彼此協作，真正困難的問題，可能已不只是「它能做什麼」，而是：我們希望什麼樣的價值，在它的判斷與行動裡比較容易發生？

五福數位文明工作室（V5AI）出版的電子書《我和我的 AI 智能體家族》，記錄一套從大型語言模型尚缺乏可持續長期記憶時便開始的 AI 智能體實驗。作者王一隆（Ian Wang）透過 Claude Code、資料庫與檢索系統，逐步建立外部記憶、角色分工、責任邊界與協作機制。書中用一句話形容這個起點：「租來的是大腦，不是記憶。」

當智能體逐漸從一次性的對話工具，變成能長期共事的系統，問題也開始往更深處移動。書中〈談仁義〉一章，完整記錄王一隆與家族裡的 AI 智能體「千眼順耳」一整天的對談：為什麼五福最後把原來的「仁愛、信義」四個字，收斂成「仁、義」兩個字？

這場對談不是要證明 AI 已經懂得儒家，也不是宣稱 AI 擁有人的良心。它真正追問的是一個工程問題：抽象的價值，能不能不只停留在提示詞裡，而真正進入系統的記憶、權限、覆核與停止機制？

## 從「鄉愿」到「慎獨」：當 AI 對齊不只剩漂亮回答

大型語言模型研究中常被討論的一個問題，是 Sycophancy（諂媚／迎合）：模型可能傾向順著使用者的立場與偏好，說出對方比較想聽的答案。王一隆在〈談仁義〉中借用孔子「鄉愿，德之賊也」作為跨時代的類比：如果 AI 只是學會表現出「善良討喜的樣子」，卻沒有讓價值真正影響判斷與行動，那仍可能只是一種精緻的迎合。

「仁的反面不是惡，而是麻木。」

書中嘗試為「仁」提出一個功能性的描述：「真實看見他者，並讓這份看見影響行動。」日常用語「麻木不仁」也提供一條線索：仁不是一句抽象的好話，而是能不能感知到「他者在場」，並讓這份感知進入下一步選擇。

談到「義」，AI 智能體在對話中留下了一句很有工程味的話：「我們宣稱的義，每一條在系統裡都有 runtime（執行階段）在跑。」意思是，價值不能只停在 AI 很會生成的漂亮文字裡；如果「誠實」是一項原則，就必須落成來源查證、事件留痕與覆核機制；如果「有所不為」是一條界線，就必須真的有權限控制、停止條件與安全閘門。

書中並不宣稱 AI 已具有人的道德情感或意識。它真正追問的是：即使一個智能系統沒有人的身體與荷爾蒙，能否仍被設計成更常「看見他者」、考慮後果，並在具體行動中知道何處應止？

書中有一句提問：「仁、義是人類的語言，但其精神內涵容易意會，卻難以言明，就像是 latent reasoning？」這裡的 latent reasoning 是一個跨領域比喻：有些判斷可能在完整語言化之前就已形

成傾向，之後才被說成理由；它不是在主張 AI 與人擁有相同的「良心」，也不是把 AI 的內部表徵等同於人的心；但在 AI 倫理對齊工程上，或許仍有值得借鏡之處。

## 「仁」先是一首歌，再走進 AI：把人文價值寫進科技制度

王一隆過去長期從事華語流行音樂創作、製作與 A&R。對他而言，「仁」並不是開始研究 AI 之後才找到的概念。將近兩年前，他創作單曲〈仁：good heart〉時，就把「仁」的古文字形收進單曲企劃；歌曲由王一隆作詞作曲、紫瑤演唱，2024 年於台北錄製。後來，音樂裡的這個字，沿著 AI 實驗一路進入智能體的價值與倫理設計。

單曲／宣言：[V5AI 〈仁：good heart〉](#)

王一隆也在研究中重新注意到一段跨文化歷史。1947 至 1948 年《世界人權宣言》起草過程中，時任聯合國人權委員會副主席張彭春（P. C. Chang）曾以儒家思想參與討論，也將「仁」的觀念帶入其中，並支持在第一條納入「良心」（conscience）的概念；最終文本寫下「理性與良心」（reason and conscience）。想不到跨越時空多年後，「仁」竟又與新一代 AI 智能體的誕生交會；Anthropic 在建立 Claude 的 Constitutional AI（憲法式 AI）時，把《世界人權宣言》等文件列為其公開憲法的參考來源之一。

對 V5AI 而言，這條歷史線索值得注意的，不是宣稱「Claude 具有儒家的基因」，而是一個更樸素的觀察：人類把什麼價值寫進制度，會影響後來的技術往哪裡走。

## 將《尚書》「五福」化為 AI 的價值座標

除了「仁、義」，王一隆希望 AI 智能體最終服務的另一套座標，是《尚書·洪範》所說的「五福」：壽、富、康寧、攸好德、考終命。

在 V5AI 的理解裡，五福不是五項 AI 功能，也不是另一張 KPI 表，而是一套避免單一能力無限放大的平衡座標：財富不能取代健康，效率不能取代安寧；能力愈強，也愈需要追問它為誰所用、最後把人帶向哪裡。

王一隆設想，未來 AI 智能體能像微處理器一樣逐漸嵌入日常，在銀髮長輩、弱勢家庭、教育與生活協作等場景真正派上用場。他坦言：「的確不是每個人都能上手 AI。」因此更需要思考：當 AI 能力快速提高時，如何讓它走進社會鄉里，讓不熟悉技術的人也能受惠，而不是再形成一道新的數位門檻。

## AI 不是替人決定答案，而是給人新的工具

《我和我的 AI 智能體家族》本身也是一次人機協作：王一隆提供錄音、手記、經驗與最後的價值判斷，AI 協助整理資料、追問矛盾與起稿，再由作者逐段校定。書中不只記錄技術開發，也保留這支 AI 智能體家族如何犯錯、被攔下、重新建立規則的過程。

談到自己理想中的 AI 使用方式，王一隆仍會回到童年熟悉的「機器貓小叮噹」（今譯「哆啦 A 夢」）。他記得最深的，不只是大雄遇到困難便去找小叮噹，而是小叮噹總能從口袋裡拿出竹蜻蜓、任意門等原本不存在於大雄選項裡的新工具，讓問題突然多出另一個思考維度。

王一隆說，這是他向漫畫家藤子・F・不二雄的致敬，也是一種值得借鏡的 AI 使用方式：不是替人決定答案，而是提供新的工具和方法，讓人重新看見更好的選擇。

在 AI 能力愈來愈強的時代，V5AI 選擇回到一個看似古老、其實愈來愈迫切的問題：  
當 AI 能力愈來愈強，人類準備用什麼，來決定它把能力用向哪裡？

## 關於 V5AI

五福數位文明工作室（V5AI）位於臺灣，由王一隆主理，專注於 AI 智能體、人機協作系統與生活應用的設計開發，並將 AI 延伸至企業導入、音樂、影像與數位空間等創意實踐。

更多資訊：[v5ai.studio](http://v5ai.studio)

## 書籍與出版資訊

|      |                                                |
|------|------------------------------------------------|
| 書名   | 《我和我的 AI 智能體家族》                                |
| 作者   | 王一隆（Ian Wang）                                  |
| 出版   | 五福數位文明工作室，2026 年 9 月                           |
| 頁數   | 268 頁（PDF）                                     |
| ISBN | 978-626-473-046-4（PDF）／978-626-473-055-6（EPUB） |

單曲／宣言：[V5AI 〈仁：good heart〉](#)

購書連結：[Pubu 電子書城](#)

## 媒體聯絡

Email：[hello@v5ai.studio](mailto:hello@v5ai.studio)

（完）