

談科論幻

悠遊於哲學、科學與科幻之間

陳瑞麟

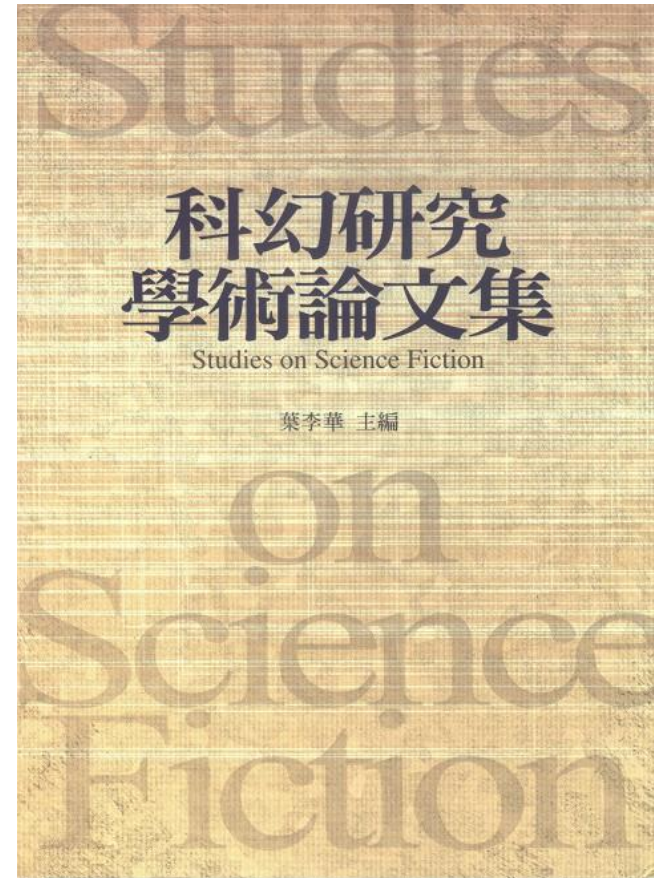
國立中正大學哲學系講座教授
國立虎尾科技大學通識教育講座

2019年03月07日

科學、哲學、科幻有什麼關聯？

- 主觀地說來，它們都是我的興趣。客觀地說來，這三者之間有內在深刻的本質關聯與歷史關聯。哲學和科學在今天似乎分屬文理，然而兩者在歷史上係出同源——希臘自然哲學，在近代也共譜一場科學革命之曲。科幻是針對未來科學的幻想或虛構，但它們常常討論「未來的哲學」。這個演講邀請大家一起來窺視科學、哲學和科幻三者的「親密關係」。

關於科幻與哲學的作品



臉書文章

- 臉書專頁：陳瑞麟的科哲絮語
- 異星世界的黯夜（上中下）—《夜幕低垂》中的科學、宗教與政治（2016-2017年間）
- 論艾西莫夫科幻中的機器人社會—兼論自由、欲求、權力與科技社會的「倫理公理」
- 未來的科技與社會學家艾西莫夫

- 科幻通論
- AI、科幻與哲學

最早的科幻？



最早的科幻？

- 格列佛遊記
- 諸島國遊記
- *A Voyage to Laputa, Balnibarbi, Luggnagg, Glubbudrib, and Japan*



最早的科幻？

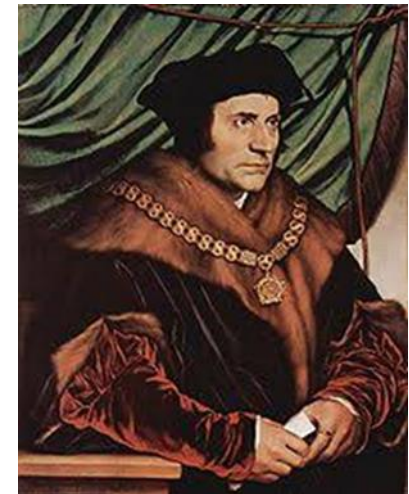
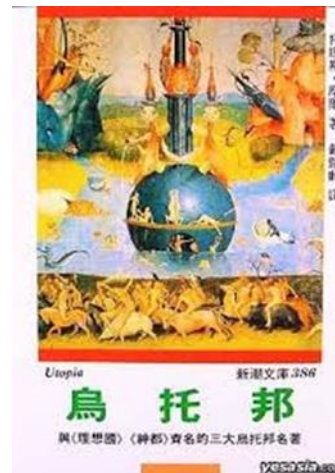
- *The New Atlantis* (1626)/ Francis Bacon



- 兩條科幻路線：
- 宇宙論玄思
- 社會烏托邦

社會（科）幻想

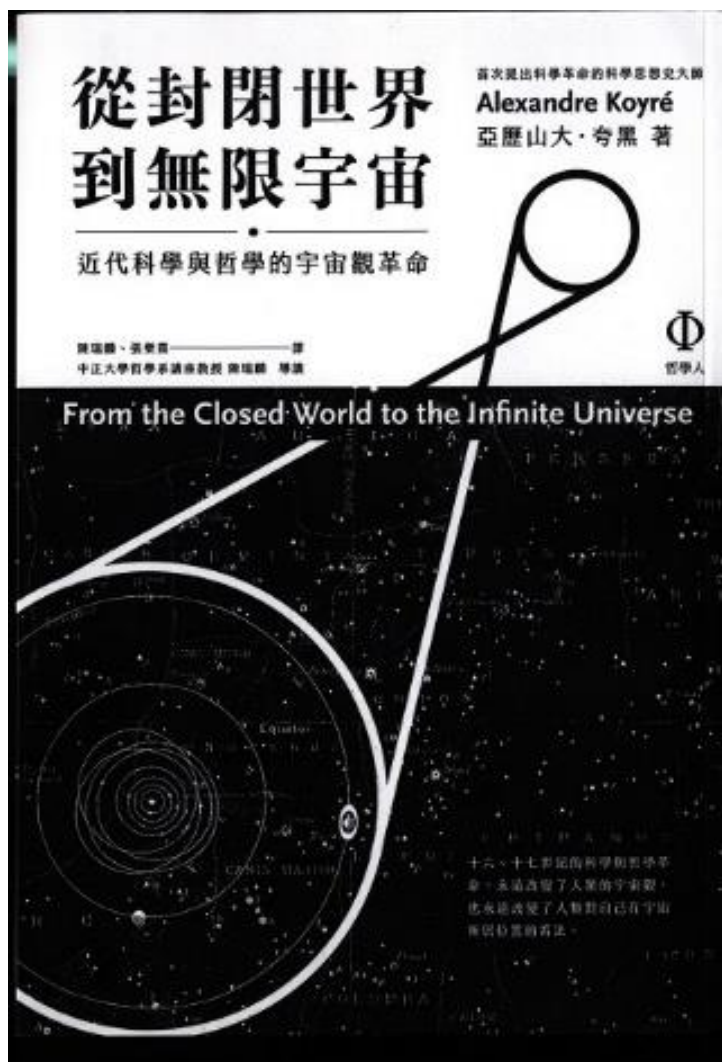
- *Utopia* (1516)/ Thomas More
- 社會科幻
- Plato: *The Republic*
- （理想國）



科學史中的未來幻想

- 宇宙論玄思
- Roger Bacon, 1220-1292的科技預言：早於 Francis Bacon, 預言望遠鏡、汽車、飛機的發明。
- Nicholas of Cusa, 1401-1464
- 大膽的宇宙論想像，背離教會正統宇宙論
- 教會正統宇宙論：地球是宇宙中心，太陽只是太陽天球的一面明鏡

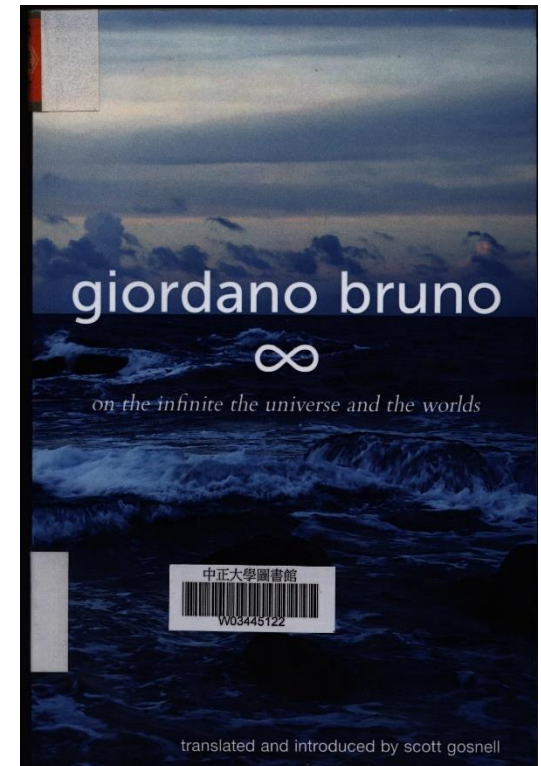
從封閉世界到無限宇宙



Nicholas 對太陽的想像

- 教會正統世界觀：地球是土水氣火構成的。地球外圍包著火。
- **Nicholas:** 「太陽體內必定有一個更中心的部分，一個「近似地球」，以及周圍繞著近似熾烈的光芒，其中有近似水的雲和清新的氣...」（言下之意，在地球外的人看地球會看到另一個太陽）

- 布魯諾（**Giordano Bruno, 1548-1600**）斷言「無限宇宙」，有無數類似地球的世界，以及無數居民（外星人）。



科幻、科學與哲學的三個基本架構： 時間向度、人性三向、心物關係

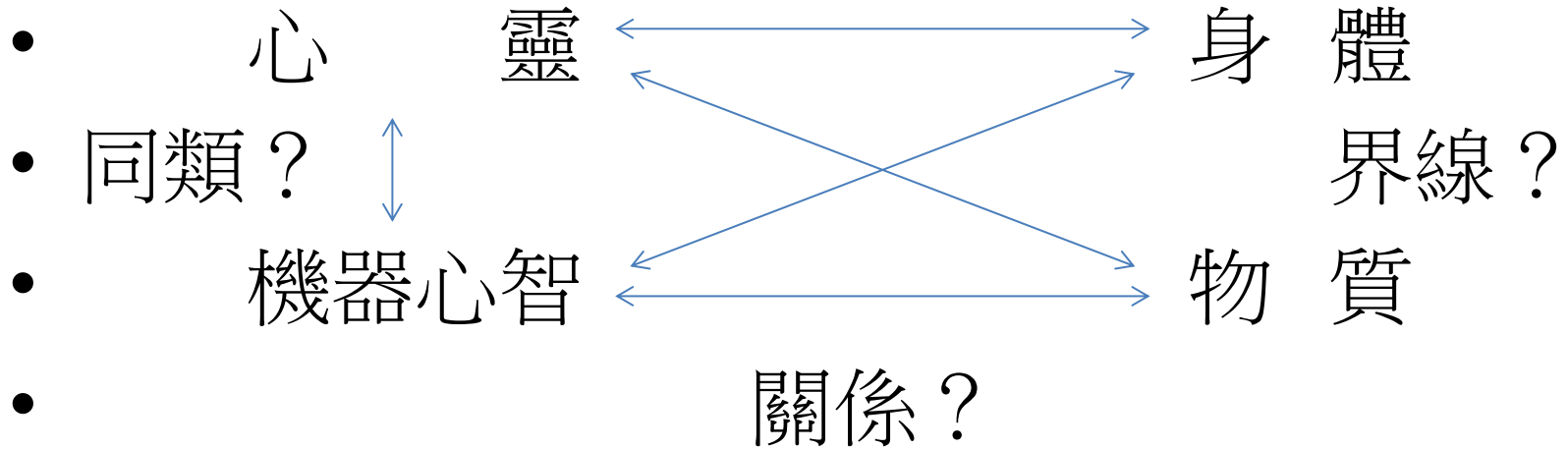
- 時間向度
- 過去—我從何處來（人類從何而來）—記憶（歷史）
- 現在—我是誰（人是什麼）—自我意識（人性）
- 未來—我將往何處去（人類將往何去）—期望（新科技）

人性三向度

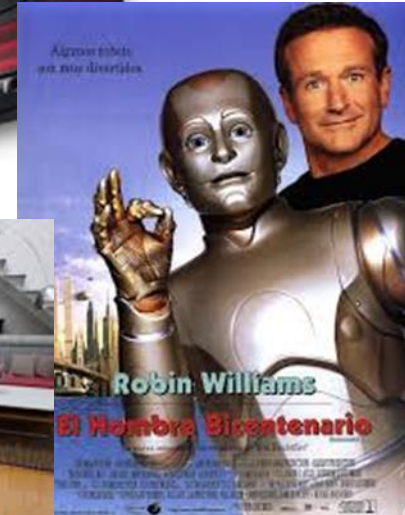
- 超人（外星人？）
 - 宗教態度
 - 倫理態度
 - 人（自我） ← 他人
 - 工具、支配態度
 - 劣人（動物、自然、機器人、人造物...）
-
- The diagram illustrates the 'Three Dimensions of Humanity' (人性三向度) as a vertical axis and a horizontal axis. The vertical axis represents a spectrum from 'Superman (Alien?)' at the top to 'Inferior Man (Animals, Nature, Robots, Artificial Objects...)' at the bottom. The horizontal axis represents a spectrum from 'Self' on the left to 'Others' on the right. The vertical axis is associated with 'Religious Attitude' (宗教態度) and 'Tool/Control Attitude' (工具、支配態度). The horizontal axis is associated with 'Ethical Attitude' (倫理態度). Blue arrows indicate the direction of the axes: an upward arrow on the left, a downward arrow on the left, and a double-headed arrow between 'Self' and 'Others'.

心物關係

- 心靈、身體、機器、心智



機器人圖片集



近來一些科幻電影



羅伯特玩假的？ 楊谷洋著



AI、哲學與科幻

- 人造人的想像歷史
- AI 的起源
- AI 的哲學議題

人造人的想像歷史

為什麼會有機器（人造）人？人形物的想像

- 想像先行，科技的進展使得想像逐漸實現...
- 人形物的想像。
- 古代中國：人俑（陶俑）（權力與宰制意志）、秦俑



Golem

- 古猶太巫術：**The Golem**（活人俑），以陶土製成的人形物，唸咒語使宛如有生命，能服從主人的命令，但若主人無力控制，也可能反過來殺害主人（科學怪人的先聲）



巫術人形

- 世界各文化都有巫術人形的想像
- 例如：日本「式神」(《神隱少女》)、
道教的紙人、巫毒娃娃…



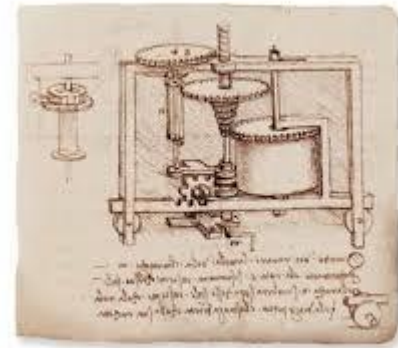
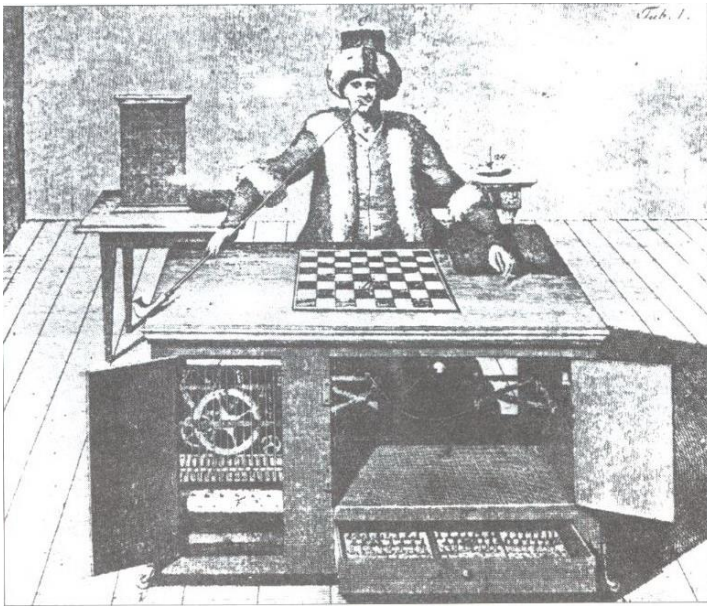
煉金術與人造人

- 十六世紀的煉金術學家帕拉塞蘇斯 (Paracelsus) 想像煉金術士能造出 Homunculus (音譯：霍爾蒙克斯)
- 日本漫畫和動畫《鋼之煉金術士》：人類對抗野心的人造人。



人類對於「智能機器」的渴望：恐怖的土耳其人

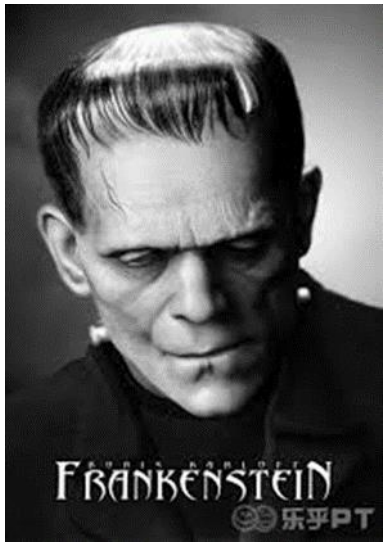
- 十七世紀之前，（西方）人類即能創造精密機器



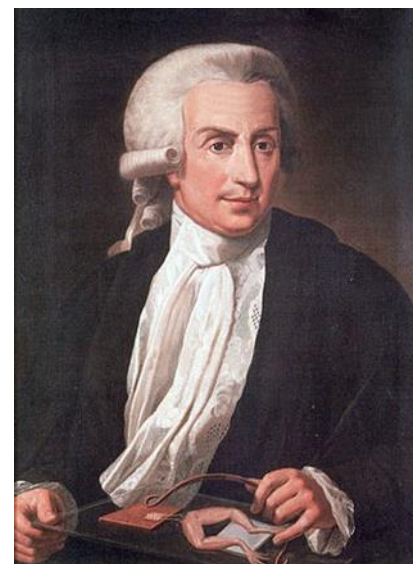
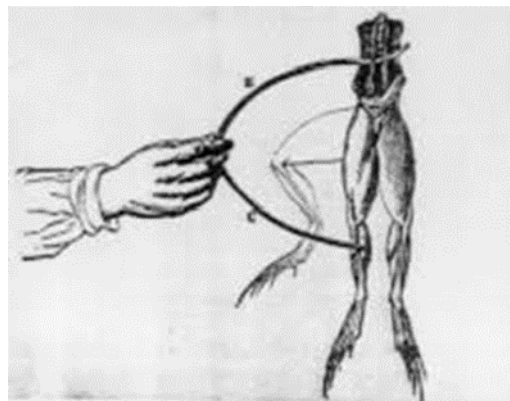
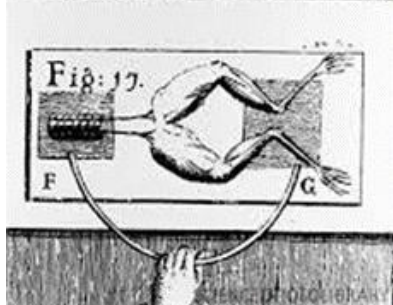
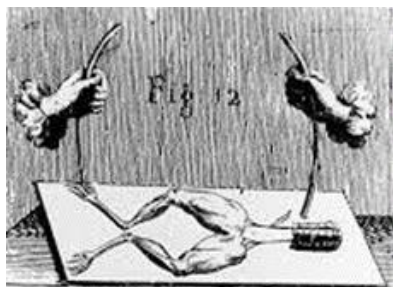
科幻

十九世紀瑪莉雪萊的科學怪人

- 科學怪人：法蘭根斯坦博士由死刑犯肢體拼湊成「人」，使用電使其有生命。



- 想像的科學背景：十九世紀解剖學的進展、十八世紀末的電學（伽凡尼的蛙腿實驗，電流刺激蛙腿使其抽蓄——想像：生命與電流有關？）...



歷史

電力和機器生命

- 十九世紀下半葉，電力機器與技術快速發展，重新喚醒機器人想像...機器化的「科學怪人」。
- 法蘭根斯坦情結(**Frangenstein's complex**)：對於「科學怪人」又愛又怕的心理，既想發展力量強大的機器僕人，又怕被傷害...

渴望人造人有長遠的歷史

- 綠野仙蹤



成人世界

-



大英雄天團

AI的起源

理性與思想的可計算與機械化

- 十九世紀末，符號邏輯被發明，思考推理可以符號化，進一步機械化（真值表、真值運算）——可以用機械來處理。
- 思考與推理變成計算？進一步用機器來計算？
- 算程(**algorithm**)與程式(**program**)：算程是一個計算的程序（根據一定的計算規則可以得到答案）；程式是一定的運作規則之組合，**run**一個特定的程式可以產生一個固定的結果。

可程式化與人工智能

- 「程式」是機器人自主移動性的來源。人類創作的程式創造了機器的「智能」（intelligence）——人工智能（智慧）。
- 智能讓機器產生自主性，使它能處理外界環境的訊息，自作「判斷」（或計算）以「決定」要如何回應。
- 挑戰：能創造出具人類自主程度的人工智能嗎？兩個難關：(1) 讓 AI 能真正「理解」人類語言意義，以及 (2) 讓 AI 能產生意識。
- 怎麼樣才算「理解」語言？
- 怎麼樣才算是有意識？

人工智能的廣義與狹義

- 廣義（野心的）的：像人一樣的人工知能。
- 狹義（現實）的：機器學習(machine learning)
- 機器學習是人工智慧的一個分支。人工智慧的研究歷史有著一條從以「推理」為重點，到以「知識」為重點，再到以「學習」為重點的自然、清晰的脈絡。(維基百科)
- 機器學習理論主要是設計和分析一些讓電腦可以自動「學習」的演算法。機器學習演算法是一類從資料中自動分析獲得規律，並利用規律對未知資料進行預測的演算法。(維基百科)

下棋的震撼

- 1997年IBM公司開發的「深藍」（Deep Blue）電腦程式系統擊敗當時的世界西洋棋冠軍卡斯帕洛夫，引發世人一陣喧嘩。
- 源於中國的圍棋卻是程式設計師無法攻克的城堡，因為它的遊戲方式與其它吃子攻殺的各類象棋、西洋棋大不相同。
- 2016年，英國 DeepMind公司（後被Google併入）開發的人工智能圍棋軟體擊敗了韓國頂尖圍棋士李世乜，又在之後的一年之間，接連擊敗所有最強的圍棋棋士，寫下人工智能嶄新的一頁。

AI 是煉金術？

- Ali Rahimi, a researcher in artificial intelligence (AI) at Google in San Francisco, California, has charged that machine learning algorithms, in which computers learn through trial and error, have **become a form of "alchemy."** Researchers, he says, do not know why some algorithms work and others don't, nor do they have rigorous criteria for choosing one AI architecture over another. *Science* 04 May 2018: Vol. 360, Issue 6388, pp. 478



AI 的哲學議題

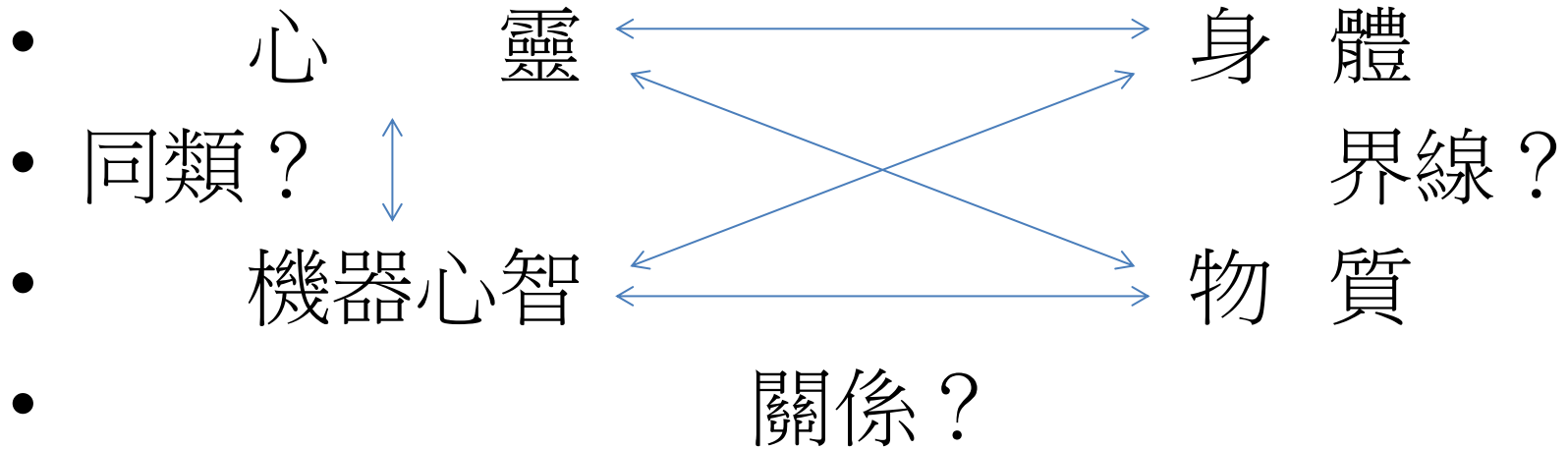
- (1) 我們和 AI 的關係是什麼嗎？
- (2) 我們如何建立研究與使用AI的規範？
- (3) 在未來，如何建立AI 與我們的社會關係？

人性三向度

- 超人（外星人？） 宇宙
- ↑ 宗教態度
- 倫理態度
- 人（自我） ↔ 他人
- ↓ 工具、支配態度
- 劣人（客體：動物、自然、人造物...）
- 人造人的位置？

心物關係

- 心靈、身體、機器、心智



AI與道德和法律：幾個科幻腳本

- 腳本一：人類與 AI 機器人的對抗。
- 腳本二：機器人自然產生道德。
- 腳本三：人類為 AI 機器人訂立道德規範。
- 腳本四：如果一個 AI 系統突然產生意識，人類應該怎麼對待它（或「他」）？有意識的 AI 是否該具備人權？人類是否該為 AI 系統訂立享有人類權利的法律？
- 腳本五：為什麼不多談 AI 幫助與服務人類的情節？

腳本一：人與 AI 的對抗

- 有意識的 AI 是演化的強者：強者會因此要統治、甚至消滅人類嗎？
- 例子：『魔鬼終結者』、『駭客任務(I)』



腳本一的延伸與變形

- 有意識的機器人可能催生賽博格（Cyborg）——人類會把自己改造成合成人，以便對抗機器人？（木城幸人的《銃夢》：肉身大腦與機械身體的賽博格漫畫）



腳本二：AI 機器人自然產生道德

- 如果機器人有了自我意識，他們會產生自利(**self-interest**)的想法，但他們也是有理性的東西。
- 理性與自利的權衡，機器人會自然產生道德意識：合作互利對所有人和機器人才是真正理性和自利的。
- 如果機器人有意識，沒有道理他們不能產生同情弱者的情感？同情共感：移情心理，把自己對自己的想法投射到他人...

哲學家霍布斯的道德起源論證（契約論）

- 理性與自利的人，會思考合作與訂立契約才是長遠有利的。不然，人類將會陷入「everyone against everyone」的自然狀態...
- 預設：人的傷害他人的能力大約平等；有思想的人都有理性，能衡量長遠利害。
- 駭客任務(II)、(III)：幫助人類的 AI（先知）。



Neo與Architect的超冗長對話場景



看看這個雞巴人！

腳本三

人類為機器人訂立倫理規範

- 我們擔憂人工智能的發展總有一天讓機器人有了意識。
- 科幻小說家艾西莫夫在五六十年前為想像中的、具有自我意識的機器人訂下了三大法則：
- 一、機器人不得傷害人類，亦不得坐視人類受到傷害。
- 二、在不違反第一法則下，機器人必須服從人類命令；
- 三、在不違反第一與第二法則下，機器人必須保護自己。
- 這三大法則能被寫入具自我意識的機器人之硬體線路中嗎？

- 機器人倫理學（robot ethics）：人類應該如何發展和使用機器人？
- 林宗德〈機器人倫理學專題報導〉
（<https://scitechvista.nat.gov.tw/c/skYJ.htm>）

科幻小說與電影中的機器人

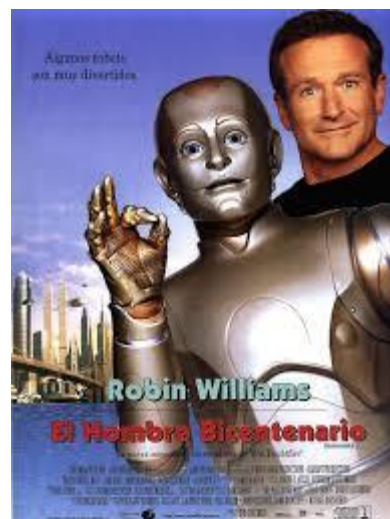
- 艾西莫夫與席維伯格，《正子人》（電影《變人》劇本之所本）
- 陳瑞麟，2006，〈正子人的人性歷程〉
- 郭文華，2009，〈與正子人共舞〉，《科學發展》11月號。
- 艾西莫夫的《我，機器人》（注意「我」字！電影《機械公敵》的故事框架是出於第六則「消失無蹤」，但徹底扭曲原著精神。）
- 艾西莫夫的機器人偵探小說系列：《鋼穴》、《裸陽》、《曙光中的機器人》（幼獅）

腳本四：人類如何對待有意識的機器人和人造人？

- 人造人的反抗
- 銀翼殺手
- 正子人

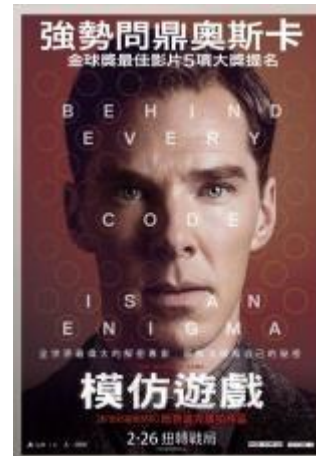


- 機器人的肉身渴望？
- 變人



AI 能擁有心靈嗎？

- 如何判斷一台電腦擁有（人性的）心靈？只是量的問題？
- 數位化的人工智能能「真正理解」人類語言的意義嗎？
- 圖靈測試（Turing Test）：讓人類同時與真人和電腦交談一段時間，如果人類無法判斷哪個對話者是真人或電腦時，就可以說電腦擁有心靈。
- 電影「模仿遊戲」：圖靈的故事。



- 「英國雷丁大學（University of Reading）系統工程學系宣布，來自俄國的尤金·古斯曼（Eugene Goostman）超級電腦在倫敦皇家學會（Royal Society）所舉辦的 Turing Test 2014 競賽首次通過「圖靈測試」，也是首次電腦騙過人類，讓人相信它是一位十三歲男孩，人工智慧歷史跨入新的里程碑。」〈電腦首次通過圖靈測試，人工智慧大突破？〉

（<https://www.ithome.com.tw/news/88507>）

哲學家 John Searle 的 思想實驗

- 中文房間論證：想像一個中文房間，有一位從未學過中文的美國人，他有一本中英文翻譯手冊。
- 這個美國人可以透過手冊和一位說中文的人對答如流，通過圖靈測試。
- 但我們能說這美國人「理解」中文嗎？

- 中文房間論證的兩個結果：圖靈測試可能不是檢驗 AI 能否擁有心靈（理解意義）的標準。
- AI 系統（例如網路上的「對話程式」）即使可以通過圖靈測試，仍然不能說它「真正理解」人類的語言意義。



■在中文屋內按翻譯規則操作的美國人可使中文屋通過圖靈測試，但他仍然不懂中文。
(本圖取自《科學的美國人》)

人工智能有可能產生「意識」嗎？

- 什麼是意識？我知道自己在想什麼、做什麼，我知道自己是一個人而不是別人（自我意識）、我知道我自己知道這些（我能意識到自己正在意識著自己的意識）...
- 如何去檢驗一台電腦有意識？
- 如何去檢驗一個人有意識？
- 電影：〈人造意識〉
（又名「機械姬」）



腳本五：感情豐富的機器人？

- 為什麼電影和漫畫總是呈現暴力化機器人的形象？事實上，真實的機器人往療癒的方向發展。所以，為什麼不多拍有感情的人與機器人互動？
- 機器人與性別：暴力化機器人和性別有關？男性傾向暴力化機器人？



機器情感的問題

- 什麼是情感？是否有機器情感？
- 如果有機器情感，它與生物的情感有什麼不同？
- 情感與感覺有何不同？情感與思想有何不同？

WALL-E



謝謝聆聽，歡迎提問

THE END