香港持續發展研究中心
香港城市大學
City University of Hong Kong

香港城市大學

香港持續發展研究中心[1]

第 28 號建議書

人工智能發展的倫理考量

卓雋傑、陳浩文、熊天佑、李芝蘭、林芬、Viktor Tuzov、李建安[2]

1. 引言

人工智能技術的政策規範討論只不過短短數年，但由於發展一日千里，全球各類機構、科技企業已制定了逾 100 份不同版本的倫理原則和指引，確保研發者及使用方都可按照獲廣泛認可的倫理標準行事。然而，如何使這些抽象的原則落實到具體的立法和規範機制，仍處於初步階段。

人工智能是香港未來發展不可或缺的一環，我們由2019年起開始透過文本研究、深入訪談以及問卷調查，辨識香港市民對人工智能在應用上的倫理原則及價值考量，並依據此提出一個可能的管治框架。

我們從問卷調查中發現，在去語境化(Decontextualized)的狀況下，香港市民對人工智能發展的 5 項倫理價值，包括: a.) 透明度； b.) 無偏差； c.) 穩健性； d.) 個人隱私； 以及 e.) 個人自由的重視程度幾乎沒有差別，都非常高。為了釐清差異，我們在問卷中加入了幾個情境(Situational)，結果「個人私隱」與「個人自由」這兩個倫理價值就變得相對突出。

我們將在本文，與各界分享這次調查結果的觀察，並據此提出適切香港人工智能發展管治框架的幾點建議。

---

[1] 香港城市大學持續發展研究中心（CSHK）成立於 2017 年 6 月，是一個開放和跨學科的研究平台，旨在促進及增強香港學術界、工業界和專業服務界; 社會及政府; 以及香港與不同區域之間的協作，並從事有影響力的應用研究範疇包括香港專業服務、一帶一路、粵港澳大灣區、綠色經濟、新冠病毒（COVID-19）等，研究項目屢獲資助，並出版多份研究報告、論文和書籍。更多資訊請瀏覽中心網頁 http://www.cityu.edu.hk/cshk 。如對本政策建議書有任何 意見，歡迎電郵至：sushkhub@cityu.edu.hk 。

[2] 卓雋傑為香港城市大學公共及國際事務學系博士生；陳浩文為香港城市大學公共及國際事務學系教授；熊天佑為香港大學經管學院特約副教授及香港持續發展研究中心國際及專業顧問；李芝蘭為香港城市大學公共及國際事務學系教授、香港持續發展研究中心總監；林芬為香港城市大學協理副校長（環球戰略）及媒體與傳播系副教授；Viktor Tuzov 為香港城市大學媒體與傳播系博士生；李建安為香港持續發展研究樞紐成員。

2. 去語境化下的倫理考量

據經濟合作暨發展組織（OECD）的定義，人工智能(Artificial Intelligence)是指一種基於機器的系統，能就人類定義的目標作出預測、提供建議，或對影響真實或虛擬環境作出決策。人工智能系統被設計成於不同的自主程度下運作。

在人工智能發展的過程中，眾多學者通過文本分析綜合研究了公共部門、私人企業和研究機構發布的多份人工智能倫理原則和指引，當中蘇黎世聯邦理工學院的研究指出，透明、正義、公平和隱私等是最廣泛獲採用，因此重要性高於其他較少被提及的價值或原則[3]。按照這些文本的結論，以及我們與 22 位業界人士、專家和學者的深入訪談，我們選定了 5 個倫理價值作為研究對象。這 5 個價值可作以下的解釋:

1) ***系統透明度***：*智能系統的運作原理及其限制是否明確，以及系統對使用的數據和信息的來源和處理過程是否有足夠的披露。*
2) ***系統無偏差***：*智能系統的運作是否存在任何形式的偏見或歧視，確保其操作公正。*
3) ***系統穩健性***：*系統的成效與安全性，包括在面對外部不穩定因素、干擾和惡意攻擊時是否能正常運作。*
4) ***個人隱私***：*系統如何保護用戶的個人隱私數據。*
5) ***個人自由***：*使用智能系統時，對用戶對行為或生活方式選擇的自由度的影響。可供選擇的範疇越廣，自由度便越大。*

在 2022 年 3 月我們以電話隨機抽樣的方式，成功訪問了 510 位年滿 18 歲的香港市民。我們在問卷中先以去語境化(Decontextualized)，也就是在未有預設情境以及解釋這些倫理原則的情況下，讓受訪者根據自己的理解來評分。

結果受訪者對 5 項倫理原則的評分相當接近，以眾數(mode)的方式來排序，全部選項得分最多的都是「7」； 改以中位數(median)的方式來排序，全部得分皆為「6」； 最後以均值(average)來排序，最低得分的「系統無偏差」(5.47)與最高分的「系統穩健性」(5.75)相差也只有 0.18。(見表一)

表一、各種倫理價值或原則以不同方式計算的得分

n=510

	系統透明度	系統無偏差	保障個人私隱	保障個人自由	系統穩健性
眾數	7	7	7	7	7
中位數	6	6	6	6	6
均值	5.61	5.47	5.71	5.56	5.75

*1 分代表「非常不重要」； 7 分代表「完全重要」

[3] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines.Nature machine intelligence,1(9), 389-399.

從上述結果反映出，在沒有特定應用技術及場景的框架和資訊引導下，受訪者普遍認同以上 5 項倫理原則在人工智能應用上的重要性，而且呈現最大化(maximization)的傾向。

3. 不同情境下的倫理考量

我們之後加入兩個不同情境讓受訪者對不同價值之間作取捨，便出現了一些變化。情境一加入「健康碼系統」作為應用例子後，認為「個人私隱」較重要的受訪者比例達 58.8%，比起選擇「系統成效」 (23.3%)較重要的高出超過 35.5 個百分點。切換至情境二「欺詐偵測系統」，選取「個人私隱」較重要的受訪者比例更達 60.4%，同樣較覺得「系統成效」較重要的高出約 35 個百分點。(見表二)

表二、不同應用情境下「個人私隱」與「系統成效」的取捨

n=510

	情境一 健康碼系統		情境二 欺詐偵測系統	
	人次	佔比	人次	佔比
(1) 個人私隱較重要	300	58.8%	307	60.4%
(2) 難以取捨	91	17.8%	71	14.0%
(3) 系統成效較重要	119	23.3%	130	25.6%
(1)與(3)差距	***--***	***35.5 個百分點***	***--***	***34.8 個百分點***

上述兩個情境中，「個人私隱」都佔了主導位置。也就是說即使人工智能發展可以幫助防止疫情蔓延；甚或協助偵測詐騙罪行，但在多數人看來，這些都不可以用「個人私隱」來作交換。

其後，我們改以「個人自由」來作對比，結果便沒有那麼一面倒。同樣在情境一「健康碼系統」中，認為「個人自由」較重要的受訪者比例，和相信「系統成效」較重要的相差收窄至 6.1 個百分點；切換至情境二「欺詐偵測系統」，選擇「個人自由」較重要與「系統成效」較重要的人數比例，差距則有 27.3 個百分點。至於新增的情境三「自動駕駛汽車」，選擇「個人自由」較重要的就比「系統成效」較重要的則多出 15.2 個百分點。(見表三)

表三、不同應用情境下「個人自由」與「系統成效」的取捨

n=510

	情境一 健康碼系統		情境二 欺詐偵測系統		情境三 自動駕駛汽車	
	人次	佔比	人次	佔比	人次	佔比
(1) 個人自由較重要	227	44.6%	284	55.8%	246	48.7%
(2) 難以取捨	86	16.9%	80	15.7%	90	17.8%
(3) 系統成效較重要	196	38.5%	145	28.5%	169	33.5%
(1)與(3)差距	--	6.1 ***個百分點***	--	***27.3*** ***個百分點***	--	***15.2*** ***個百分點***

我們在此也可作一個有趣的反向解讀，那就是覺得在某些情況下「個人自由」不及「系統成效」重要的情境排列，依次為「健康碼系統」(38.5%)、「自動駕駛汽車」(33.5%)以及「欺詐偵測系統」(28.5%)。出現這個結果，其中一種比較合理的解讀，是這些情境是否牽涉個人和別人的生命安全。

此次問卷的調查時間正值疫情期間，「健康碼系統」被視為是防止疫情蔓延的重要手段，因此相當高比例的人或許都願意放棄「個人自由」(出入指定處所花時間去掃瞄，以及在沒有滿足特定條件下－例如接種疫苗－被拒入場)來讓健康碼系統發揮成效。事實上，我們也曾在問卷中問及，在「自動駕駛汽車」的情境下，如果面對只有兩種選擇：「撞死行人」或「轉向撞上隧道牆壁撞死自己」，結果發現寧願撞死自己的受訪者比例達 51.4%，而說會撞死行人的只有 30.7%。

另一點值得留意的是，綜合表二分析，即使在情境一「健康碼系統」下有 38.5% 受訪者接受「系統成效」比「個人自由」重要，但卻較少人(23.3%)願意接受「系統成效」凌駕「個人私隱」。這除了顯示出「個人私隱」的受重視程度，或也反映出大眾相信，當局可透過系統設置技術保護私隱。

4. 道德困境與出路

在不同的價值取捨背後，受訪者對於在個別情景應否採用人工智能技術意見相當分歧。當中在應否採用「健康碼」以及「欺詐偵測」兩個情景中，贊成及反對的聲音相差只有 2.2 個百分點至 3.3 個百分點。至於在「自動駕駛汽車」的情景中，反對的受訪者比例佔了絕大多數。(見表四)

表四、不同情境的應用意見?

n=510

	應否採用?	
	應該	不應該
情境一・健康碼系統	49.9%	47.7%
情境二・欺詐偵測系統	45.6%	48.9%
情境三・自動駕駛汽車	29.4%	65.4%

表四的調查結果與當時的社會討論氛圍相當吻合。在 2020 年 8 月推出「健康碼」的時候，支持者力陳這是控制疫情蔓延的有效方式，並可藉識別不同風險人士，以解除其他過多的社交距離限制，助力社會經濟復甦。但反對一方卻擔心「健康碼」會附設行蹤紀錄及讀取個人資料，變相侵反了個人隱及自由。可以預見，隨著人工智能技術深入發展，運算精準度要求的提高，不同應用情景的資料存取、活動實時監控以至是軟硬件設計等都會日益精細化，以上的爭議也只會愈來愈尖銳。

我們在問卷中就不同情景可能出現的道德困境作提問， 結果發現無論是哪一個應用情下，一旦出現道德困境，受訪者都希望「受影響市民」的意見會是最優先被考慮。(見表五)

表五、出現道德困境應優先考慮誰的意見?

n=510

	情境一 健康碼系統	情境二 欺詐偵測系統	情境三 自動駕駛汽車
1. 公共機構	19.4%	24.6%	8.8%
2. 私人企業	2.0%	1.5%	1.7%
3. 第三方專業人士	12.0%	14.5%	24.5%
4. 受影響市民	59.9%	53.5%	58.4%

值得留意的是，受訪者對於「第三方專業人士」的認可度也相當高，認為他們的意見應先被考慮的百分比，達一成二至近兩成半，當中對於「自動駕駛汽車」這相對較新穎的技術，信任度尤高。相反，受訪者對於技術應用服務的最大提供方「私人企業」卻缺乏信心，在各種情景下認為私企的意見應該最先被考慮的只有區區的 1.5%至 2.0%。

5. 小結

當我們與社會不同持分者討論人工智能發展的倫理原則及政策規範時，不少人的即時反應都認為這些只是「抽象」以及「空泛」的概念，因此很難制定出可操作的具體框架。然而透過上述的問卷調查分析，我們認為出現「抽象」或「空泛」的主因，或許只是大家往往沒有將倫理原則與現實情況具體結合起來開展討論。

我們的問卷調查結果顯示出，在抽空的去語境化(Decontextualized)下，所有倫理價值都被視為是重要的，無論是系統的成效以至個人的自由及私隱。但諷刺的是，恰恰是當所有選項都是同等重要及不可逾越時，那麼社會及政策制訂者將感到無所適從。因為在多數情況下，每項社會政策皆要作出不同程度取捨。

因此倫理價值原則要在政策制訂中發揮規範作用，那麼倫理價值就要被應用到不同的具體情境中。這正是我們在問卷中，加入情境(Situational)選項的原因。結果也正好反映了，在加入不同應用情境後，之前重要性被視為幾乎無差異的各項倫理價值，就有了判斷取捨的可能性。

此外，當在制定政策時面臨道德困境，我們毫不意外地發現受訪者認為受影響的主體，會是最重要的被諮詢對象。然而調查也反映出，有相當多比例的人相信專業第三方人士，亦是重要的溝通渠道，可以協助理順人工智能發展過程中面臨的難題。

我們相信以上的調查結果及觀察，都有助於社會制定適合本地人工智能發展的監管和規範措施，我們將在下一篇文章中提出幾點建議。