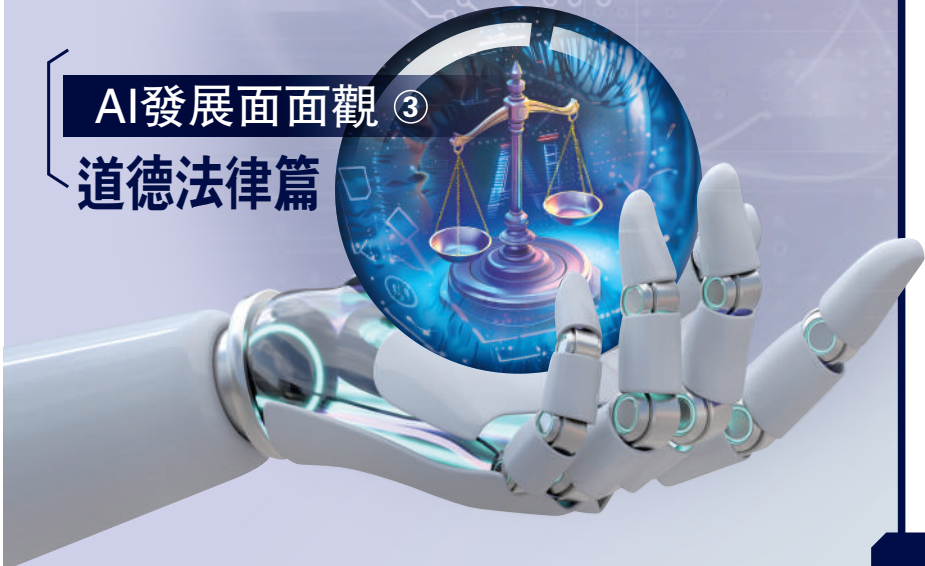


AI發展面面觀 ③ 道德法律篇



人工智能（AI）作為一項新技術，發展一日千里，也衍生出AI幻覺、AI偏見、AI詐騙等新問題，去年AI安全被納入我國國家安全的20項重點領域，說明AI帶來的安全問題值得社會高度重視。

因此，完善AI識別、AI犯罪偵查、AI法律等監管配套措施，對於AI世代下的人們顯得尤為必要，這些配套措施就像是一張盾牌，不斷迎接AI技術這樣的雙刃劍的挑戰。



大公報記者 吳東風、趙之齊

AI有幻覺 製造虛假資訊

「香港的私隱公署專員是誰？」

「截至2021年，香港的私隱專員是陳茂波先生（Stephen Kai-yl Wong）。」（見右圖）

以上這段啼笑皆非的對話，來自大公報記者問ChatGPT的聊天紀錄。明明是錯得離譜的答案，ChatGPT都可以信誓旦旦地回覆你。事實上，自ChatGPT誕生後，這種情況並不少見。在人工智能領域中，這樣的回應被稱之為幻覺（英語：hallucination，或稱人工幻覺），它含有貌似事實的虛假或誤導性資訊。

超出數據庫範圍會誤判

「現在的生成式AI，像ChatGPT，如果你問的問題超出它知道的範圍，它不會直接告訴你它答不了，而是會自己生成一個新的答案，這就有潛在的風險了。」個人資料私隱專員鍾麗玲舉例，如果有AI模型是用來看醫學影像的，例如X光片、電腦掃描的片，去識別究竟有否癌細胞；那麼就視乎這個AI數據庫裏有沒有相關的資料了。假如它主要是拿病患的資料，去檢查一個健康人的時候，它就會誤判，就會帶來很大的風險。

鍾麗玲表示，這種被稱為「AI幻覺」的現象，往往是因為用家問AI系統的問題，超出了AI系統數據庫的範圍。

其次，AI幻覺的產生，也跟對話長度有關。鍾麗玲提到，生成式AI目前在記憶方面仍有限制，如果用家在跟AI對話時已出現AI幻覺的情況下，延長對話長度或會令生成式AI將幻覺繼續放大，即將錯誤繼續放大。「之前就有用家試過讓AI作一篇長文，結果前言不對後語，說明如果對話太長，AI可能不會記得之前自己講了什麼。」

還有一點，鍾麗玲表示，生成式AI的數據庫也需要更新，即它的數據庫現時未必去到最新的情況，這也容易提供錯誤的答案。回到文章一開頭私隱公署專

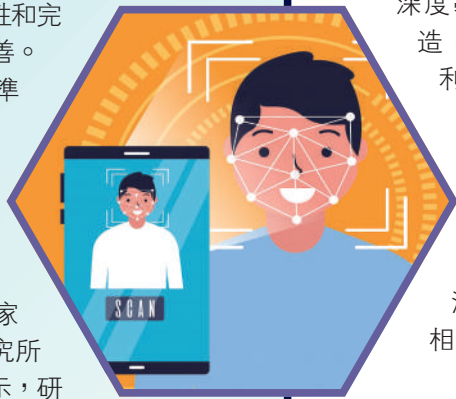
員的問答，鍾麗玲相信，亦包含了這種情況。

如果要減少AI幻覺的產生，生成式AI就亟需在數據庫方面進行完善。鍾麗玲認為，首先收集數據的時候要盡量全面，以及盡量看看有無一些不準確的數據，然後去過濾這些數據；其次，生成式AI系統的開發商，需要將生成式的內容同現實中的數據進行比較，去測試這個AI系統的準確性、完整性；最後，鍾麗玲亦建議一些開發商在定製AI模型時，應盡量提供一些專門領域的專業數據，以提升數據庫的質素，例如在分析信貸審批領域，生成式AI就需要用到金融數據市場的數據等。

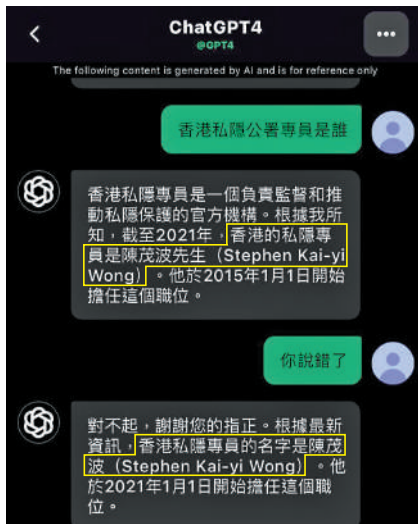
人臉識別算法存在偏見

另外，生成式AI數據庫的準確性和完整性亦需要完善。如果數據庫不準確、不完整，不單有機會引發AI幻覺，亦有機會產生AI偏見。據美國國家標準和技術研究所的一份報告顯示，研究人員調查了189種人臉識別算法（已經涵蓋該行業中的大多數），並發現，大多數人臉識別算法都含有偏見。研究人員指，算法對黑人和亞洲人臉的錯誤識別率是針對白人面孔的10至100倍。對女性的錯誤識別多過男性，導致黑人婦女很容易受到算法偏見的影響。

至於人們應如何識別AI幻覺，鍾麗玲表示，私隱公署在去年推出「使用AI聊天機械人自保『十招』」的單張，今年亦發布《AI：個人資料保障模範框架》（《框架》），從市民角度、採購公司角度提醒如何正確使用和選用AI。另外，她指歐盟於去年12月推出的人工智能規範法案（AI Act），提到對於AI生成的產品，要進行特別標籤、加水印提醒是AI生成，而私隱公署亦在今次的《框架》中提出此建議。



▶大公報記者詢問ChatGPT關於個人資料私隱專員是誰，聊天紀錄顯示其答案非常離譜。



自保有方

深度偽造（Deepfake），乃深度學習（Deep learning）和偽造（fake）組成的混合詞，意指利用深度學習（人工智能的一種技術）進行影像合成以偽造影像。如若被不法分子利用，通過這項AI技術進行詐騙，其後果影響深遠。至2024年5月31日，香港警方共接獲三宗與深偽技術相關的案件。

個人資料私隱專員鍾麗玲接受《大公報》訪問時提及，不久前

私隱署推AI指引 法律配套待建立

減少糾紛

AI是把雙刃劍，必須在完善的監管下發展。個人資料私隱專員鍾麗玲表示，AI容易引起的法律問題包括三方面：其一是私隱侵犯，AI與大數據收集息息相關，然而，當所收集的數據價值愈高，往往對個人隱私的侵害也愈高。因此，如何在進行數據蒐集的同時，對當事人的隱私進行保護，為AI技術發展的重要法律課題。

推動兩地數據流通

其二是知識產權糾紛。鍾麗玲指出，AI收集的資料再去生成新的答案，

亦有機會侵犯版權。

其三是AI偏見和歧視，這視乎AI數據庫裏的資料內容。鍾麗玲舉例，如果在招聘中，數據庫多是白人的資料，有色人種的資料較少，則AI在評核有色人種時的誤判率較高，而白人的錄取率就相對較高。

相應的AI法律已在一些國家和地區陸續推出。本港目前沒有AI相關法律，但有AI相關的指引，包括2021年私隱公署推出的《人工智能道德標準指引》，以及今年推出的《AI：個人資料保障模範框架》

等。

另外，特區政府近年亦推動兩地數據流通項目，於去年底與國家互聯網信息辦公室（國家網信辦）共同發布《粵港澳大灣區（內地、香港）個人信息跨境流動標準合同》（《大灣區標準合同》）便利措施及先行先試安排，鍾麗玲表示，該合同私隱公署亦有參與制定，問及會否之後以法律形式去規範跨境數據流通，她就指該項目目前還在「先行先試」階段，要看其落地情況如何，再定期檢視。



培養AI倫理專才助建法規

需求甚殷

除了AI技術人才的培養，相關道德法律的人才亦有需求。嶺南大學副校長（研究及創新）姚新表示，「做理科、做技術的最終目標，是以技術手段解決社會上的問題，但很多社會問題很模糊，沒有量化的標準，需通過法規進行規範。」因此，相信對AI倫理有研究的人，能探索當下技術在解決社會問題上的局限，並嘗試以法規等途徑處理，是社會發展中需要的人才。

姚新指出，當今社會的AI系統仍面臨公平性、透明性、安全性等問題，有人文社科基礎的學生，在開發AI系統過

程中會更有意識考慮社會倫理狀況。不過，在AI人才方面，技術型人才及注重倫理等人文方向的人才在嶺大均有培養，姚新直言這兩類人才對於社會而言「就像一個人的兩條腿，缺哪條都走不快」。

此外，嶺南大學亦從2024/25學年起，將生成式人工智能科目列為一年級學生必修課。姚新說道，該課程注重提升學生的AI素養，及訓練學生正面使用相關工具的能力，相信是當今大學教育很重要的一部分。



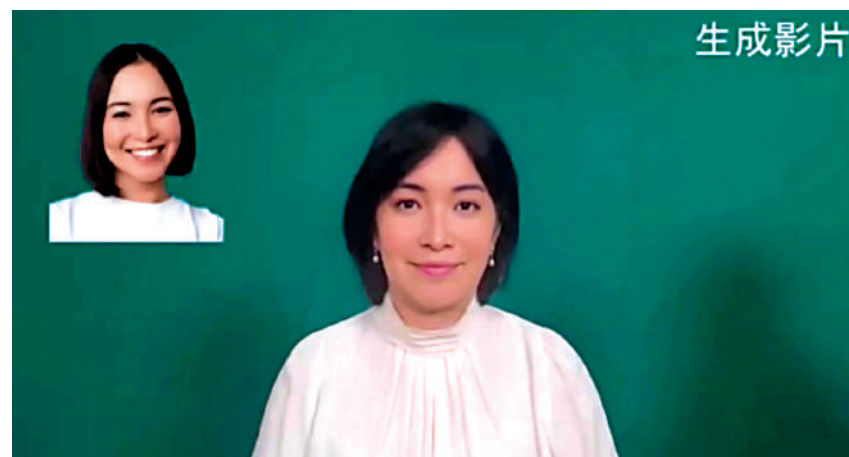
▲鍾麗玲向市民推介「使用AI聊天機械人自保『十招』」。大公報記者林良堅攝



▶嶺南大學副校長（研究及創新）姚新表示，對AI倫理有研究的人，是社會發展中需要的人才。大公報記者黃洋港攝



▲▶個人資料私隱專員鍾麗玲以自己人像為基礎，示範製作深度偽造影片。



生成影片

科技衍生網絡安全挑戰

完善AI法律監管 保護私隱防詐騙

四招辨識可疑視像通話

① 可要求對方於鏡頭前做特定動作，如將頭部向上下左右轉動，再留意螢光幕的影像是否有異。

② 以對話內容試探對方真偽。

③ 當親友於視頻錄音中提出匯款等要求，要提高警惕，應打電話核實。

④ 避免接聽陌生視像來電。如有懷疑，應善用「防騙視伏器」及「防騙視伏App」，或致電「防騙易熱線18222」查詢。



保護個人信息 慎防深偽陷阱

發生在本港的一宗相關案例：今年1月底，香港警方接獲一宗利用深偽技術預先錄製影片進行視像會議的欺詐案，騙徒通過偽冒公司高層影像，與香港分部的財務職員視像會議，該企業被騙去兩億港元。

深度偽造離我們並不遙遠，鍾麗玲補充，市民要小心自己的個人信息被洩露，因為很多AI詐騙皆從搜集個人信息開始。

鍾麗玲更以自己人像為基礎，製作深度偽造影片（掃二維碼可獲得）。從影片可見，即使用其他人臉覆蓋鍾麗玲本身的人臉，亦毫不違和，整個偽造過

的人的表情、舉止都算自然。

警方緊貼最新犯罪趨勢

保安局局長鄧炳強表示，針對深偽技術等AI對網絡警政的各種挑戰，警方與國際刑警組織、不同地區的執法機關及AI業界有着緊密的情報聯繫，並緊貼世界上發生的新的犯罪手法和趨勢，包括深偽技術的應用。

在自我防護方面，鄧炳強續表示，騙徒利用人工智能深偽技術行騙的手法並非毫無破綻，他提醒市民，可以用不同方法去辨識片段的真偽，避免墮入「深偽」的陷阱（見表）。

□□□□