《AI 大講堂》
第三集 - AI 安全與政府角色：保障你我數碼生活
視像文字稿

歡迎收看《AI 大講堂》
我是方健儀
人工智能為大家帶來不少便利
但同時亦有一些新挑戰
因為不法分子
會利用 AI（人工智能）
進行深度偽造
即是 Deepfake（深偽技術）
他們會進行詐騙
又或者發放假資訊
令市民蒙受精神或金錢上的損失
所以在 AI 大時代
大家要學懂如何保護自己
相信這是每人都要學習的一個議題
我們這個節目《AI 大講堂》的目的
是透過一系列的專家對談
用最簡單的方法
讓大家學會 AI 和善用 AI
而這集的嘉賓就是數字政策辦公室
副數字政策專員（數據治理）
吳秉宗博士
作為 AI 治理專家
他要和我們分享一下
究竟有甚麼 AI 常見的陷阱
以及我們應如何好好保障個人私隱
吳博士你好
健儀，你好
剛才提到有些不法分子
會進行深度偽造
他們可以換臉、變聲
亦都可以發放一些詐騙
或者是釣魚訊息
聽起來也防不勝防
所以這些資訊或者內容

對普羅市民有甚麼影響
以及為甚麼 AI 生成的內容
這麼容易得到人的信任？
AI 當然可以幫助我們展現創意
但其實有些不法份子
真的可以用一些相片或者一些聲音
去模仿其他人的身份
從而作出一些詐騙的行為
健儀
我今天帶了一條錄音想考考你
好，先聽一聽
Akina
我知道你有做電台節目主持
對聲音很熟悉
我現在想問問你
你能否分辨這條錄音是由 AI 生成
還是我真人的錄音？
我就這樣聽，吳博士
很難分辨
因為我今天才第一次與你見面
但我有做好資料搜集
我有在網上聽過你的聲音
剛剛那把聲音與真人聲音對照的話
十分相似
但你問我的話
應該也是偽造聲音來的，對嗎？
對，其實這段聲音是生成的
除了聲音之外
其實也可以生成一些影像
我今天帶了另外一條影片過來
由「個人資料私隱專員公署」提供
我們看看
大家可以看到
我們可以將一張相片
套在一個真人上
而真人在移動的時候
樣貌亦能與相片上的人相似
其實就很容易取信於市民身上

從而作出一些詐騙的行為
所以大家可以想想
如果一些 AI 技術
一些深偽技術
被一些不法分子利用的時候
其實影響很大
其實兩年前已經有些案件涉及詐騙
有些跨國公司的視像會議的高層
其實是用深偽技術偽造出來
涉及的金額高達兩億元
所以你想一下是有很多深遠的影響
聽起來十分令人關注
所以我們面對這些 AI 製造出來
或者生成的內容
大家都會問
究竟我們怎樣分辨哪些是真
哪些是假的？
其實要分真假越來越難
幾年前大概有六成的機會
你會分到是真是假
但因為技術繼續發展
現在做到一些高質素的影像時
可能只剩下四分一機會
甚至乎有些人說
我用 AI 去分析 AI 會不會可以？
其實是可以的
不過也只有一半機會
所以你可以想到
就是沒有一個完美的方法去分辨
深偽技術出來的影像或者聲音
其實市民都可以好好裝備自己
我今日帶了兩個小提示過來
第一個就是「停一停」
如果你收到一些要求
例如叫你過錢、交出密碼時
要「停一停」
第二就是「想一想」
舉個例子，你可能會問多一點

對方跟你一起認知的事情
看看它是真是假
甚至乎有些人家裡有一個暗號
可以知道是真是假
亦都可以掛電話
然後再打給對方
甚至乎上一些官方的網站
查一下是真是假
要用的方法有很多種
但最重要是要小心
同時在過程中亦可以留意到
深偽技術的影像有時會有少少缺陷
舉個例子，你動一下頭
動一下手的時候
可能會有一些不尋常的抖動
這些就是值得留意的地方
市民如果懷疑受騙
可以致電香港警務處
反詐騙協調中心
「防騙易 18222」熱線
或直接報警
要細微觀察和想一個家庭的暗號
這個很重要
除了去看 AI 生成的內容之外
其實現在越來越多人
喜歡跟 AI 聊天
例如有些人會將自己的個人資料
給一些 chatbots（聊天機器人）
即是一些 AI 聊天機器人
除此之外，也有一些朋友
用 AI Agent（AI 代理）
即是 AI 代理去處理日常事務
連密碼也給它們
我就想，我們正在使用 AI 的時候
如何去保管好自己的個人資料？
有沒有一些個人資料
是一定不可以給 AI？
這個問題非常好

其實一些很個人的敏感資料
是不可以給 AI 的
很多生成式工具
或剛才說的代理式人工智能
可能會涉及一些安全隱患
包括一些權限過高或者資料外洩
所以有些很敏感的資料
是不應該給予的
權限亦都要檢查清楚會否太高
舉個例子，在外國有些安全的研究
曾經發現有一個聊天機器人
它已經有五千萬用戶
但後台做得不好
導致一些資料外洩
有三億條訊息外洩了出來
裡面包括一些個人資料、健康資訊
甚至一些財務資料
都是非常不理想的
另一個案例就是外國有些企業
有些同事將一些代碼放在大模型上
甚至將一些機密的會議記錄放上去
令到有些商業機密外洩
其實市民可以
用一個「紅綠燈」模式
去想想甚麼資料可以交出去
甚麼資料不可以
市民可以用「紅綠燈」的思維
紅燈「絕不交出」
例如你的身份證號碼、護照號碼
銀行帳戶、信用卡號碼
登入密碼、OTP 驗證碼
醫療記錄、健康數據
黃燈的話就要「想清楚先」
例如個人姓名、聲音、影片
公司名稱
因為這些資料
傳輸到 AI 平台的時候
可能會讓 AI 生成個人的頭像或訊息

這些影像亦可能會用來訓練 AI 模型
綠燈是「可以使用」
完全無個人訊息或敏感訊息
無法辨認身份的資料
最後給一點實用建議大家
在用這個平台的時候
大家要問問自己
「我信不信任這個平台
我的資料放上平台的時候
如果外洩會怎樣？
還有沒有其他渠道
可以代替這個方法？」
這個「紅綠燈」挺好用
大家要謹記
其實除了個人層面之外
公司層面都經常用到這個生成式 AI
例如有些公司或者工作間
未有一些很清晰的 AI 政策的時候
究竟第一步應該怎樣做？
以及應該怎樣提醒員工
在用 AI 工具的時候要加倍小心？
這也是一個很好的問題
在公司層面
如果大家能夠好好利用 AI
是可以提升很多工作效率的
個人方面亦都是
我們看到一個研究說
如果一個同事他能夠用好 AI
其實他相比他的同行
可以有更加高的收入
甚至多百分之五十以上的收入
當然，這不是簡單地問 AI 一些問題
其實要將人類獨有的能力
與 AI 緊密結合
舉例，情商、創意、身體的靈活
道德判斷力和責任感
和 AI 緊密結合
就可以做到剛才所說的事

至於在工作期間
尤其是在企業方面
其實我們可以有三個步驟
舉例，制定一些內部 AI 使用政策
有一些敏感的資料
舉例，如客戶的訊息、個人的訊息
一定不可以放出去
第二部分，就是一些培訓和提醒
除了培訓員工之外
亦都要提醒他們
要覆核再覆核自己的 AI 生成的結果
是否真的合乎用法
另外如果有些事情真的不幸發生了
怎樣有一個上報的通報機制
最後有一個風險評估
大家可以參考數字辦公室的
《香港生成式人工智能技術
及應用指引》
那裡也可以做一個很好的起點
我們休息一下
稍後回來繼續和吳博士聊天
休息回來，繼續《AI 大講堂》節目
我們繼續和數字政策辦公室
副數字政策專員（數據治理）
吳秉宗博士聊天
吳博士你好
我有一個想法
其實除了個人層面也好
或者是公司層面也好
都要面對一個「秤」
這個「秤」就是說
我們現在很鼓勵大家
要擁抱這個創新
但是另一方面
我們怎樣平衡風險控制這個層面？
非常好，我又帶了一張照片過來
這張照片是剛剛發明汽車的時候
剛剛發明汽車的時候

很多馬匹在路上行走
當時的人就想
如果車撞到馬會怎麼辦
於是他們就派了一個人
拿著一支紅旗在車前面走
讓車不會撞到馬
大家都會想到
車最快的速度就變了是人的速度
扼殺了車的能力
之後人類就定一些規則或指引
不再需要拿紅旗的人
車亦可以行快一點
慢慢地馬匹被淘汰
路上全部都是車
這亦是平衡創新和風險時
要留意的地方
用 AI 的時候其實
已經不是一個「用不用」的問題
而是怎樣可以安全和負責任地用
有些人見到 AI 的時候
可能有不同的看法
舉個例子，有些人很擁抱 AI
他一見到工具就馬上用
完全不考慮風險
其實是有些問題的
有一類人見到 AI 就立即抗拒
完全不想用
因為很擔心
這亦影響了他個人能力
或經驗上的提升
最務實的方法就是取得平衡
一步一步地做
例如在個人的層面上
你可以分析工作風險的高和低
將一些比較低的風險
可以嘗試去做
舉例，搜集資料、整理資料
這些都是比較低風險的

而在企業層面
可以制定一些簡單的指引
跟同事說，一定不可以將敏感訊息
客戶訊息放上去人工智能的平台上
所以有一個中庸之道是很重要的
不如說一下政府的角色
因為在這個 AI 治理的框架裡面
我們想看看在香港政府方面的部署
是怎樣去保障市民的權益
吳博士本身是來自數字政策辦公室
你們的角色又是怎樣？
其實 AI 不能只靠政府
也要和市民及機構共同推進
政府的角色不是幫大家決定
用不用 AI 工具
而是一個守門人的角色
政府會訂立一些原則，出一些指引
提醒一些風險
及推進創新的時候
又可以用一個安全
負責任的方法去推進
政府在 AI 治理上有三大支柱
第一是「原則」
第二是「指引」
第三是「應用」
第一個「原則」上
我們有《人工智能道德框架》
大家可以想像一間餐廳的衞生守則
無論你煮甚麼菜都要遵守
第二個就是《香港生成式
人工智能技術及應用指引》
就好像餐廳裡廚師的操作指引
無論他拿到甚麼工具
他都要留意一些步驟
舉個例子，會不會有 AI 的幻覺
會不會有資料保安的問題、偏見等
其實這個就適合開發者
服務提供商和市民去參考

第三就是「應用」上，有一個
《政府部門人工智能應用指南》
就好像餐廳的標準作業程序
無論 AI 炒出來的菜是怎樣
那個同事都要檢測一下、監測一下
好像 AI 生成的東西出來
同事要核對
這個比喻很形象化
大家很容易吸收到
其實我聽吳博士的講解之後
我發現我們要建立市民
長遠對 AI 的信任
這個很重要
所以我們下一步的關鍵是怎樣？
我們怎樣在和國際接軌之餘
建立一個可以是很有香港特色
亦適合我們自己的
一個 AI 的治理框架？
現在 AI 最重要的是信任
我們剛才說的治理框架
下一步最重要的是
落實一個可操作、可驗收的機制
這個就是「香港人工智能研發院」
未來的其中一個主要角色
研發院在 AI 治理方面有三件事
第一件事是制定一些 AI 標準原則
亦都可以與國際接軌
第二件事，是提供一些 AI 安全評估
令到用的時候，安心一點
第三，提供跨界別 AI 合作平台
推動整個業界的賦能及發展
政府亦積極與
如 OECD（經濟合作暨發展組織）
這些組織商討 AI 治理情況
一方面接軌國際的標準
另一方面也要配合
內地的 AI 治理框架
走出一條符合香港情況

亦都可以兼顧創新與安全的路
最後，吳博士能否給大家
一些 AI 小錦囊
當作總結一下
大家用 AI 的時候會更得心應手
其實 AI 已經無辦法避免
用到 AI 的時候
對我們的工作效率都會大大提升
當然 AI 會帶來一些風險
但相信市民是有能力可以去避免
首先第一是
記得要小心保管個人資料
任何敏感訊息、健康資訊
財務資訊都要小心保管
第二是「停一停，想一想」
第三是遇到一些可疑事情的時候
或者可疑電話的時候
我們真的要提高警惕
可以向對方確認一下他的身份
看看他是否真人
其實真人是不怕問的
只有騙徒才怕你問多幾句
吳博士，我呼應你
因為我是真人來的
所以我要多問一點
例如現在我們經常都很關心長者
又或者是一些小朋友
他們去面對 AI
例如有些人打電話給他
或者用生成式內容的時候
有沒有一些針對性的建議給他們？
例如長者
好，長者的話
可以提議他們
小心任何這些類型的電話
其實涉及一些錢
或者「你中獎了」
或者有一份新的包裹

不知道去哪裡領取
或者你犯了某些事
其實要多多提醒他們
任何事情，也要找家人去商量一下
如果騙徒假冒家人的話
其實應該早早和長者
做一些家庭的暗號
只有你們兩個知道
令長者安心
因為長者的防範意識比較弱
所以要多多靠家人在身邊的提醒
真的，我見過有些長者
收到假冒是兒子的電話
亦都是深信不疑
然後就過了錢給他們
所以在這些情況下
他們應該怎樣應對？
例如要多問幾條問題
還是要多點懷疑
令到騙徒無所遁形
首先我們真的要提醒長者
剛才說的「停一停，想一想」
不要馬上聽到一些要過錢
很急接著收線的事就馬上去做
先「停一停」
不要提供任何東西
不要提供任何敏感資訊
「想一想」的時候
不只是核實對方身份
甚至要找家人談一談
看看是否真實的事情才去做
另外小朋友方面
因為我看到很多小朋友都機不離手
他們面對 AI 生成的內容的時候
可能片段很短
又或者他們要思考的時候
究竟怎樣分真與假
貼士的錦囊又是怎樣？

其實現在很多 AI 生成影像的平台
生成出來的影像
大概是十秒左右
所以你會看到一些很短的影片
但是很精彩的
也可以做到一些很有創意的事
但你想想一個小朋友
如果你給他看五分鐘
他在一些串流平台上看這些影片
其實他可以收集到很多這樣的資訊
這些資訊有些當然是有創意的
但亦都有一些比較危險
舉例，有一些影片是一些貓、狗
從一些很高的地方跳下去
小朋友會覺得原來可以這樣
會不會我也可以這樣做？
或者原來有些更加誇張
甚至乎人都可以跳上跳下的
那些就要非常之小心
希望家長可以多陪伴小朋友
甚至在觀看的時候幫他們選擇
告訴他們甚麼是生成的
甚麼是不合理的
這些片段通常只有幾秒鐘的時間
我們大人也要花一點時間去思考
對於小朋友是不是也是一個挑戰？
絕對是，小朋友的思維沒那麼快
我本身也有幾個小朋友
都是年紀很小的
他們收到訊息的時候
真的可能以為是真的
他們未必可以這麼快
就已經處理到究竟是真是假
因為可能已經播放到下一條影片
同一時間
是接收不到這麼多數據量
有些情況是不合理的
一看便知道是 AI 生成

亦都不應該相信
甚至家長可以幫他們選擇
一些比較安全的渠道、內容
給他們看
感謝吳博士的分享
吳博士剛剛跟我們分享
AI 實用的小貼士
以及大家如何使用得更加安心
這個節目《AI 大講堂》
目的是透過一系列的專家對談
讓大家認識到在 AI 不同發展的領域
最重要大家能學會 AI 和善用 AI
歡迎大家繼續關注