人工智慧概論

CH11:AI在資訊安全的應用

National Taiwan Ocean University
Dept. Computer Science and Engineering

Prof. Chien-Fu Cheng

11-1 資訊安全面面觀

- 物聯網資安
  - 物聯網可分成三層架構為「物件感知層」、「網路傳輸層」、與「應用服務層」。
    - 物件感知層：在物件感知層的資安，要慎防被植入殭屍程式。
    - 網路傳輸層：運用AI 分析偵測負載內文有異常時，可予以預警，同時也包含阻斷攻擊的偵測。
    - 應用服務層：應用服務層的攻擊行為最為多樣化，所以各項行為都應該運用AI 進行分析，並對攻擊行為提出警示 (Alert)。

11-1 資訊安全面面觀

圖**11-2** IoT 各層的資訊安全。

11-1 資訊安全面面觀

- 資安產品或服務
  - 資訊安全產品或服務就是為防範企業資訊資產遭受各種安全威脅，以確保資訊資產之機密性 (Confidentiality)、完整性 (Integrity) 與可用性 (Availability)。
    - 內容攻擊是資安常見手法之一，電子郵件和網站又是重中之重。
    - 威脅管理又分成內部威脅以及進階威脅管理。
    - 認證與存取管理都是資訊安全的基礎。有了AI，身分辨識便有更新穎的做法：運用人臉辨識、語音辨識、虹膜辨識等人因模式進行認證。
    - 其它如弱點評估、資安事件管理等。

11-1 資訊安全面面觀

圖11-3 資安產品／服務類型。

11-1 資訊安全面面觀

- 以場域劃分的資訊安全
  - 人員進出管制：
    - 針對特殊場域或門禁進行管理。
    - E.g. 門禁卡
  - 實體裝置的資安：
    - 在實體裝置或設備的互動下之資訊安全。
    - E.g. ATM設備、IP CAM
  - 場域安全：
    - 通常可能涉及公共安全事項。油、水、電和核能電廠等關鍵基礎建設統計，是一件國安的議題。越來越多產業的基礎建設有監視控制與資料擷取系統 (Supervisory Control And Data Acquisition, SCADA)，尤其以製造業為最多。

11-1 資訊安全面面觀

- 數位服務系統的資安所涵蓋範圍有：
  - 雲端資安防護：
    - 主動防護機制，有效回應網路威脅。
  - 數位監控防護：
    - 建立數位鑑識資料庫、網路異常行為偵測、網路防護虛擬化技術。
  - 密碼安全：
    - 加解密、軟硬體憑證。
  - 鑑識及檢測鑑識：
    - 弱點偵測掃描、資安檢測技術及資安鑑識分析。
  - 適應性安全架構：
    - 整合預測、防禦阻隔、偵測及回應。

11-2 AI 發展下的資訊安全案例

- 深度學習下的深度偽造
  - Deepfake 除了拿來「換臉」之外，現在也拿來「換聲」了。
  - Deepfake 有兩個階段，第一階段為訓練階段；第二階段為測試階段。
  - Deepfake 也可以針對聲音造假，Deepfake 是針對被合成者的聲紋進行合成。然而不只有聲紋，還包含語氣和語速。

圖11-4 Deepfake 換臉程序。

11-2 AI 發展下的資訊安全案例

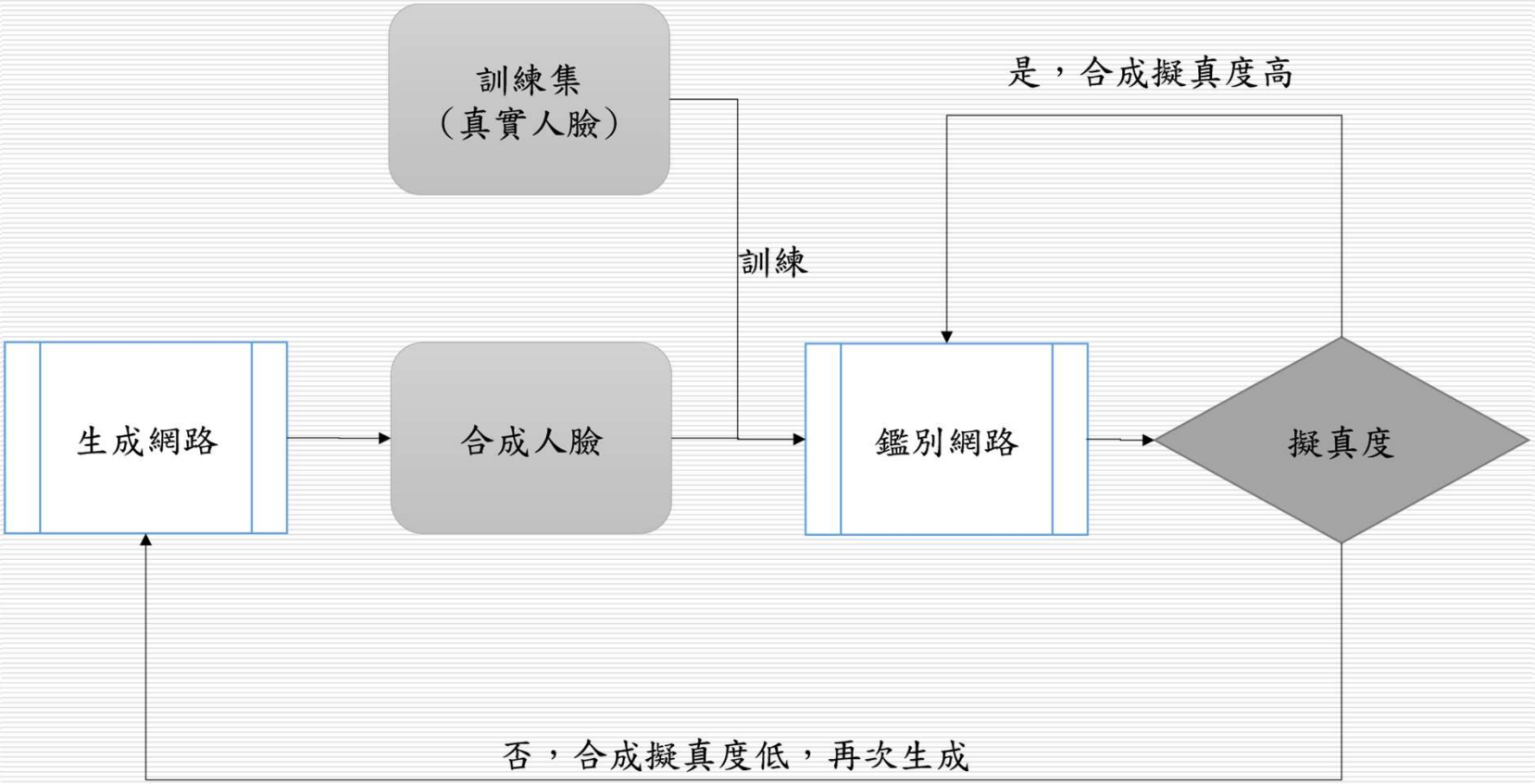

圖**11-5** Deepfake 的人臉合成。

11-2 AI 發展下的資訊安全案例

- 當你的臉不再屬於你自己 | DEEPFAKE換臉技術【厚厚解說】09:55min
  - https://www.youtube.com/watch?v=WgpMz7PEZnc
- Deepfake深偽技術解密! 人工智能換臉助長假新聞! 怎樣分辨演算法偽造的影片? 05:07min
  - https://www.youtube.com/watch?v=YvSYGetM9_4

圖11-6 用Deepfake 定位出人臉特徵點進行偽造表情。(圖片來源：cbn.globoradio官網。)

11-2 AI 發展下的資訊安全案例

- 不過語音的深度偽造有比較多的限制：
  - 第一是背景如果太多雜訊，偽造效果較差。
  - 其次是偽造的語音通常要分段，容易被識破。
  - 最後是偽造的語音無法即時轉換，所以通常不能即時通話。
- Facebook 也進行偵測與封鎖Deepfake 影片。散布假影片也違反相關法律規範，不可不小心。

11-2 AI 發展下的資訊安全案例

- 用AI 揪出無所不在的駭客
  - 名為Darktrace 的資安服務能偵測雲端、物聯網、互聯網(Internet)，甚至電力線網路的資安漏洞，並且提出警告。
  - Darktrace: Protecting the Modern Worker 01:47min
    - https://www.youtube.com/watch?v=ecv1nAOv0z0

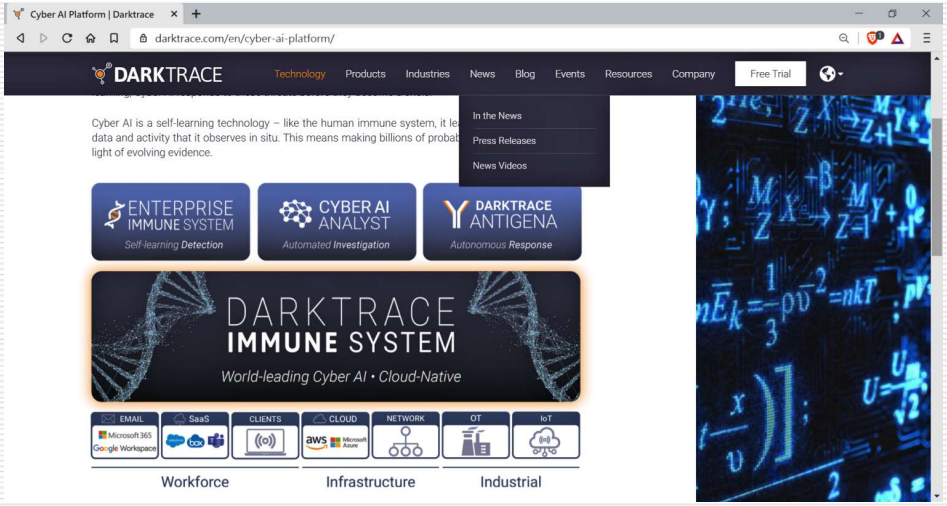

圖11-7 Darktrace 的資安平台。(圖片來源：Darktrace 官網。)

11-2 AI 發展下的資訊安全案例

- 人工智慧強化郵件伺服器
  - 電子郵件伺服器的管理上，有一個很基本但很重要的功能就是垃圾信件的管理。
  - 而人工智慧可以運用強大的演算法來讓系統經由大量數據進行垃圾郵件的判讀。
  - 要識別和防止垃圾郵件，可以結合不同的特徵判定：
    - 郵件信譽評價
    - 客製化IP 評等
    - 複合式垃圾郵件防護
    - 寄件者標題與內容自然語言比對
    - 其它等等

11-2 AI 發展下的資訊安全案例

- 最特別的是，如何揪出這些詐欺信件，最重要的一樣AI 技術是「寫作風格DNA 判別」的AI 技術。
- 用寫作風格DNA 判別的AI 技術，捕捉信件內容並分析行為及意圖。

11-2 AI 發展下的資訊安全案例

- 其它的資安AI 應用
  - AI 在資安上就是一種攻防：
    - 非營利機構Netsafe開發了一種AI 機器人，叫Re:scam，這個AI 機器人就運用來反詐騙。
    - 還有一種AI 攻防是用AI 破壞AI 學習，在語音中插入噪音，讓聊天機器人總是進行錯誤的學習。所以，在微軟推出聊天機器人Tay 不到24 小時就下架。
    - Deepfake 問世以來，許多人在非惡意下也用AI 換臉做了很多影片。美國矽谷的史丹佛實驗研究院 (SRI International) 的AI 中心就發現，Deepfake 的假影片中，人物眨眼都會稍有不自然的狀況。

11-2 AI 發展下的資訊安全案例

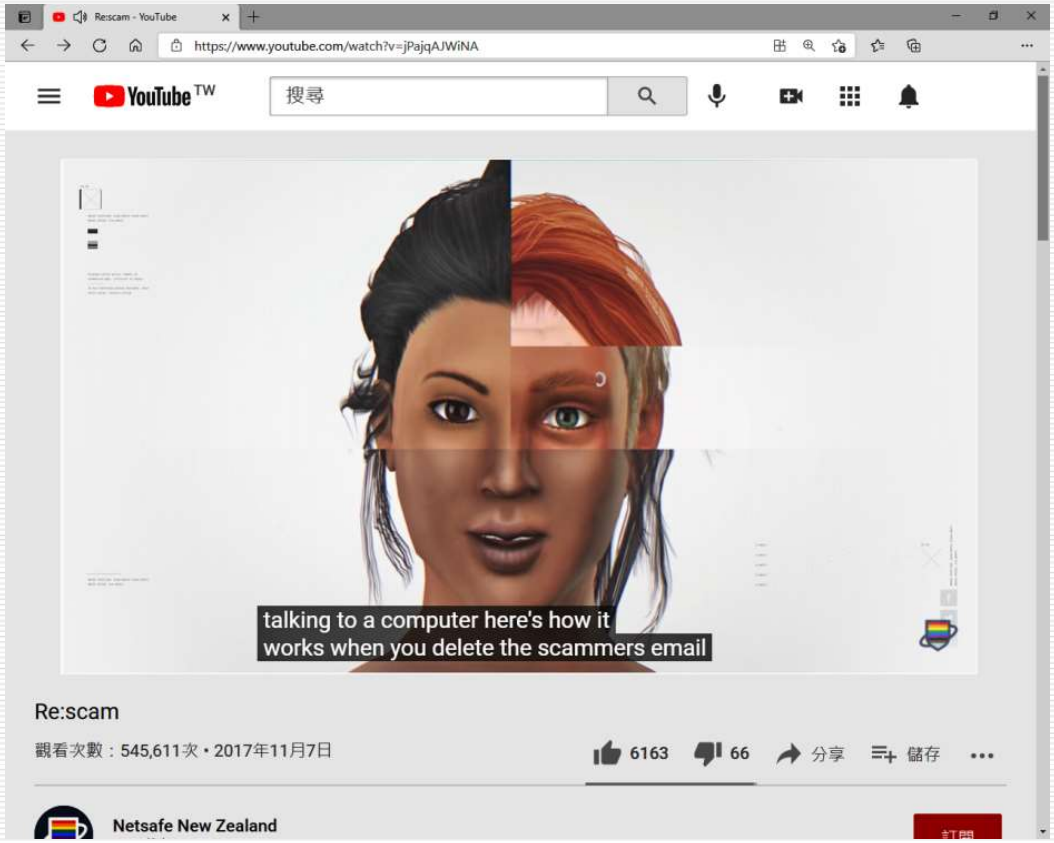

圖**11-8** 反詐騙AI Bot Re:scam。(圖片來源：TWITTER 官網。)

11-2 AI 發展下的資訊安全案例

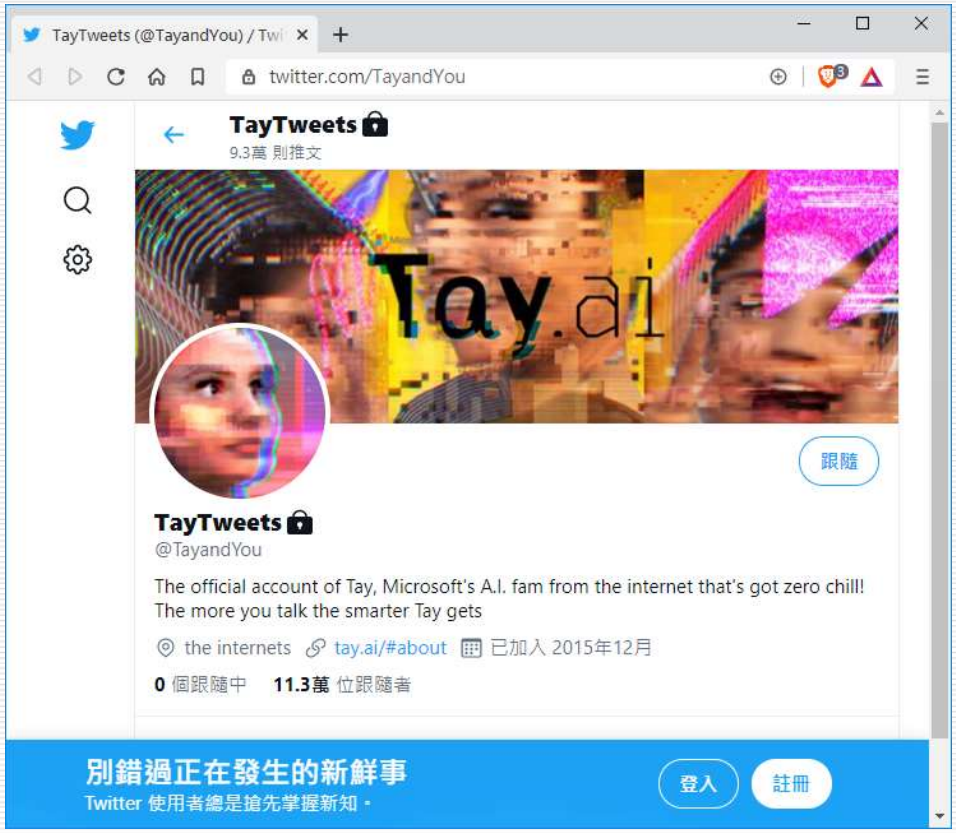

圖**11-9** Tay 的Twitter。(圖片來源：Twitter。)

11-3 正視方能避險久安

- 有人運用AI 進行資安攻擊，研究人員與AI 資安專家在運用AI 進行反制。
- 未來在資安的攻防戰中，其實是一個全民運動。

Sources

- 投影片資料來源說明：
  - 本投影片之內容出自於書商所提供之投影片，並根據實際授課需求進行補充及修改。