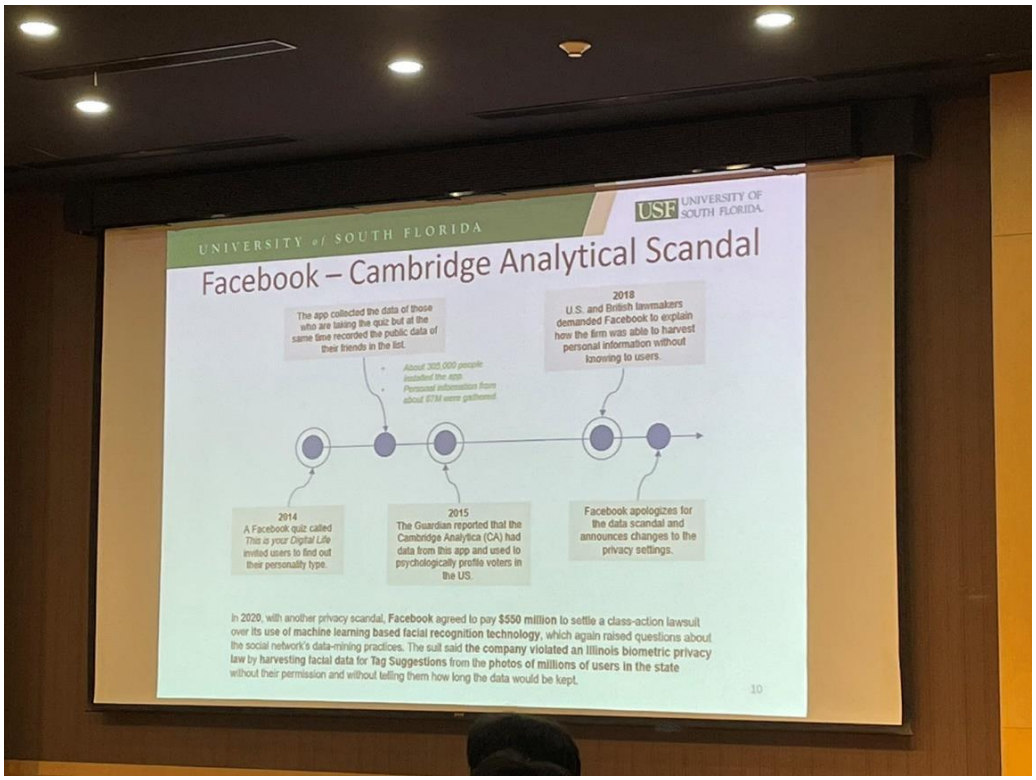
國立臺北大學 112 年度　電機資訊學院

活動名稱	電資院大師系列講座 主題：Privacy-preserving machine learning 講師：University of South Florida professor (台大電機系客座教授) 張致恩先生
活動日期	112 年 03 月 17 日星期五
活動時間	13:30-15:00
活動地點	電資院 B1 萬榮講堂
活動內容	這次的主題跟機器學習的資訊安全有關，講者講了很多研究的經驗談和一些團隊研究的技術層面和解決方法。 首先介紹背景，AI 大數據的領域介紹、Privacy-preserving 在機器學習領域的應用。為什麼講者要做這樣的計劃，注重現在的網路隱私，想辦法解決這樣的問題，接著介紹最基礎的機器學習步驟，目前運用在哪些地方，信用卡交易、指紋、臉部辨識，data 有生物特徵、健康資料、財務資料、公司顧客群資料，這些都牽扯到隱私的問題，講者把他提出來討論。目前的現狀我們訓練時會用免費的 data 當作我們的訓練 data，這些是免費的，我們常常會在生活中無形的被 google,Facebook 等大公司搜集個人的隱私資料。machine learning as services 是現在的趨勢，但其中也會面臨到許多隱私的問題，其中也提到目前除了攻擊 data set 的隱私之王外，也有可能攻擊 data set 的正確性。 講者也介紹了一些他們團隊搞個資的問題和發生的故事，例如臉書的隱私意外流出的問題，講者也把 time line 和他們演講的開始講的很清楚，也把重點拉回到 social network 的 data mining 的 privacy 問題，法律和資訊安全也開始注意到這一塊。google ,apple 也發展和意識抬頭，要告訴客戶他們有解答來保護用戶的隱私，例如 RAPPOR(randomized aggregatable privacy preserving ordinal response)，我們用他的 web browser 他們會抓到我們可能會用的，把他抓出來，另外蘋果會搜集我們常用的字、詞彙和表情，把我們這些資料加一點馬賽克上去，把原始資料做一些修改，因為他們最終的目的是要做統計分析，不用這麼清晰的 raw data。 Privacy related threats 為題目 ，回到資訊安全上，機器學習中也有許多資訊安全上的問題，機器學習上也有幾種 attack ，我們需要注意的，保護隱私我們可以用 adding random noise 的方式來將資料打馬賽克，這是一個方法，例如 randomized response 來用隨機的方式得到可能正確或可能錯誤的問卷回答。機器學習越用越廣，我們已經把很多資料和模型放在雲端上面，我們如何去保護個人資料是非常重要的議題。 另外，講者也提到，2011 年講者就開始做資訊安全的計劃，講者也播放了之前被美國電視台訪問的影片，講說密碼有可能會被 typing 的習慣給駭客竊取走，講到說要確認你是不是用戶，接著做應用軟體病毒掃描的技術，最後做的就是用手機上做的人工智慧，現在很多到需要在大的電腦，手機的小的嵌入式裝置是現在開始做人工智慧的趨勢，不同公司的作法都不太一樣。

活動成效	講者花了很大的篇幅講解現在科技上機器學習上用到的許多大數據潛藏著很多隱私權和被駭客攻擊的議題，很多時候我們只是和親朋好友出去玩，發文打卡，就會被臉書或是其它社群平台記住我們生活周遭的人臉，資訊等等，很多時候都是無意的把個人資料給這些公司，講者也相信目前被報出來的事件也只是剛剛開始，我們真的要多注意。 另外講者也呼籲大家，在學校裡面可以吸收多一點不一樣的知識，很多時候我們會在不知道或是意外的時候發現我們所研究的東西是很有用的，或是很有前瞻性、未來性，像講者的博士學生就被重金抓去臉書做資訊安全的部分。 現在很多公司提出的 privacy concern raised 技術，講者要我們思考，他們提出這樣的技術，是因為他們找到 solution ，那他們沒有提出的，是不是就有可能是根本沒有找到 solution ，其實我們的資料還是十分赤裸的晾在網際網路間。 講者也有介紹新的書，電子版已經出來了，書也陸陸續續的在各大學的圖書館都有上架，我們可以去看。講者團隊的程式會放在書上就會放在 GitHub 上，也呼籲大家不要一昧的相信 chat GPT 寫的程式，他很有可能就是 GPT 去 GitHub 上找的答案，我們還是要懂一點程式的邏輯。 我們也要去思考目前資訊安全遇到了哪些問題，除了生活上常見的問題，軍事上也可能遇到衛星定位的影像資訊被偷等等。目前在機器學習上的攻擊有 de anonymization attack (re identification)、模型的參數也可以被找出來(parameter inference attack)；模型也可以做 model inversion attack，目標找資料去訓練；membership inference attacks 他可以找出來你的資料有找哪些出處，可以去惡意改變那裏的資料（例如 chat GPT 的訓練資料被破壞，那大家使用上就會出現很大的問題)。

網路上最大的事情是網路霸凌，別人會網路肉搜、挖你的隱私，沒辦法做到 The right to be let alone ，這是 100 年前美國羅賓漢（Brandeis)律師提出的概念，已經超出保護財產，而是一個更新穎的隱私概念，現在美國就有新的演習計劃（DAPPA Brandies program)

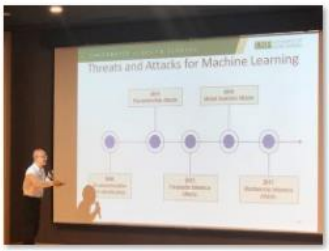
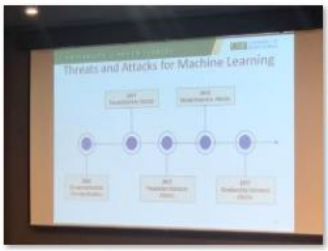
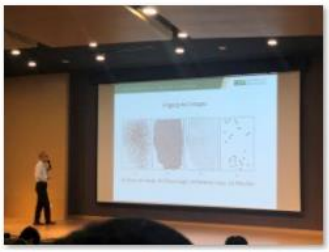
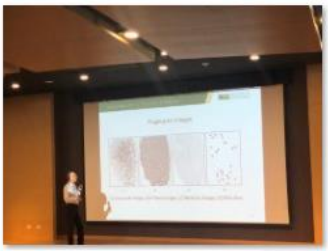
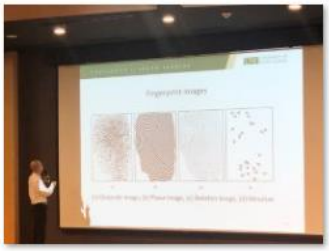
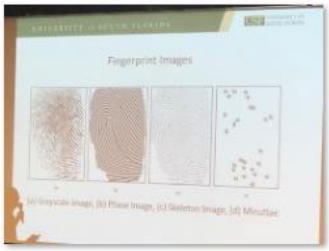
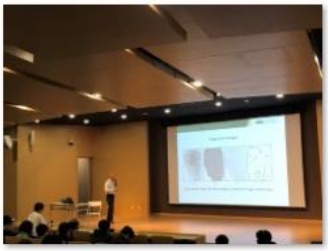
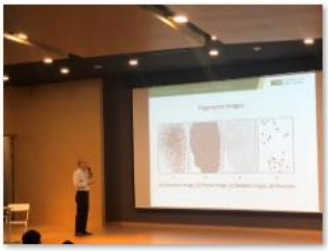


臉書的 Cambridge analytical scandal 事件，2014 年臉書給用戶做問卷，連結臉書本來的用戶資訊，婚姻狀態、聯絡什麼樣的朋友，結果被其他公司的人偷走賣給其他客戶，臉書創辦人祖克柏被叫去立法院解釋，如何搜集 personal information ，臉書也必須開始思考要如何保護用戶資料的問題，新聞媒體也將此事也報的很大，大家也開始意識到這樣類似的問題有很大的嚴重性。

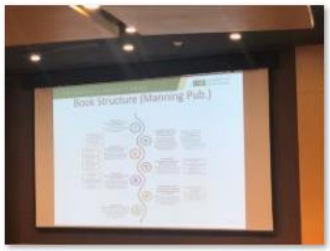
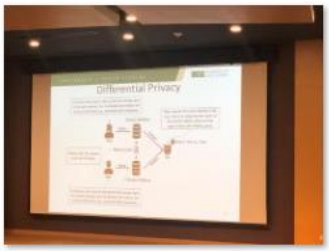


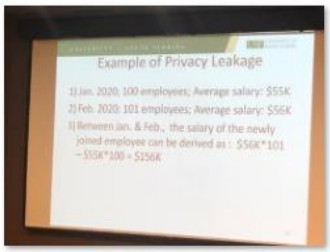


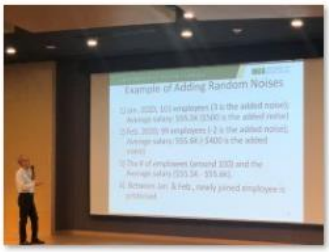




Threats and Attacks for Machine Learning

Threats and Attacks for Machine Learning




Fingerprint Images



Good Practices

Good Practices

Book Structure (Manning Pub.)

Differential Privacy

Example of Salary Database

Example of Privacy Leakage
1) Jan. 2020: 100 employees; Average salary: $55K
2) Feb. 2020: 101 employees; Average salary: $56K
3) Between Jan. & Feb., the salary of the newly joined employee can be derived as : $56K*101 − $55K*100 = $156K



Example of Adding Random Noises






Privacy-Preserving Machine Learning
USF

Privacy-Preserving Machine Learning
USF


Machine Learning (2/2)
1) Classification applications:
2) The data used can be

國立臺北大學 NTPU
電機資訊學院
College of Electrical Engineering and Computer Science

前瞻科技研究中心
Center for Advanced Technology Research

國立臺北大學
National Taipei University

112電機資訊學院
大師系列講座

112/03/17 Fri.
13:30-15:00

Privacy-preserving machine learning

張致恩 博士

美國南佛羅里達大學電機工程學系教授、臺大電機系客座教授、
Associate Editor-in-Chief of IEEE IT Professional、
Handling editor of Journal of Microprocessors and Microsystems

112年高教深耕計畫補助

主辦單位：國立臺北大學電機資訊學院、前瞻科技研究中心
協辦單位：資工系、通訊系、電機系