

# 【講座報名】我們可以信任人工智慧嗎？

## 講座介紹

人工智慧 ( Artificial Intelligence ) 在近來已逐漸應用在我們的各生活層面中，舉凡金融、交通、教育、政治、醫療、警察執法等，皆有人工智慧的應用實例。我們正將各領域中過去需由「人」來處理的專業判斷，逐漸移轉給人工智慧，進行自動化的決策。

而在人工智慧逐漸廣泛應用之際，也衍生了一系列的問題。其中最核心的問題之一是：人工智慧所做出的決策是可被信任的嗎？

例如，面對「有」標準答案的問題 ( 如圖像、語音、文字辨識等 )，人工智慧系統的可信與否，可能是取決於其能否不受各種精細的欺騙、誤導、干擾，精準做出符合當下外在世界事實的判斷；而面對「仍無」標準答案的問題 ( 如涉及「價值判斷」或「對未來的預測」的問題 )，則在什麼意義下，人工智慧所做出的判斷，又可說是被信任的？

這場講座將從技術研究者的角度，討論人工智慧若可被信任，在技術上要面對哪些挑戰，研究者又要如何建立起一個能對抗那些挑戰的系統，使人工智慧的判斷結果最終可被信任。

※本場講座之簡報為英文，但演講將以中文進行。

## 時間

1 月 22 日 15:00-17:00

## 地點

台北市身心障礙服務中心 6F 集會室 ( 台北市中山區長安西路 5 巷 2 號 )

## 主持

莊庭瑞 副研究員 ( 中研院資訊所 )

## 講者

陳品諭 博士 ( IBM Research )

## 講者介紹

陳品諭博士 ( Dr. Chen Pin-Yu ) 。 Research staff member at IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. He is also the chief scientist of [RPI-IBM AI Research Collaboration](#) and PI of ongoing MIT-IBM Watson AI Lab projects. Dr. Chen received his Ph.D. degree in electrical

engineering and computer science and M.A. degree in Statistics from the University of Michigan, Ann Arbor, USA, in 2016.

Dr. Chen's recent research is on adversarial machine learning and robustness of neural networks. His long-term research vision is building trustworthy machine learning systems. He has published more than 20 papers on trustworthy machine learning at major AI and machine learning conferences and has co-organized workshops on adversarial learning for machine learning and data mining such as KDD' 19. His research interest also includes graph and network data analytics and their applications to data mining, machine learning, signal processing, and cyber security. He was the recipient of the Chia-Lun Lo Fellowship from the University of Michigan Ann Arbor. He received the NIPS 2017 Best Reviewer Award, and was also the recipient of the IEEE GLOBECOM 2010 GOLD Best Paper Award. Dr. Chen is currently on the editorial board of PLOS ONE.

**活動日期:**

2020/01/22 -

15:00 至 17:00

**地點:**

台北市身心障礙服務中心 6 樓