

AI的偏見與歧視 甘偵蓉從ChatGPT談起

學習新視界

【記者黃國暉淡水校園報導】資圖系邀請國立清華大學人文社會AI應用與發展研究中心博士後研究學者甘偵蓉，於5月23日在B302A會議室做人文社會科學與產業實務創新鏈結計畫專題演講，主講：「AI的偏見與歧視—從ChatGPT談起」，約50位師生與會。

甘偵蓉從ChatGPT為例，闡述AI所帶來的影響及改變，不只在產業應用，而是涉及生活各個層面，相關倫理討論是否可以更多元以及人社學者如何更積極參與AI發展及回應挑戰。

資圖三簡郁婷演講後表示：「老師舉例ChatGPT本身就有一些不同立場，會生成帶有特定政治色彩的話題，有偏見的，也可能不公平。所以，科技雖然儘量客觀，但其實仍有影響。」



Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification

Joy Buolamwini, Tessie Debra Proceedings of the 1st Conference on Fairness, Accountability and Transparency, FAccT 2019, 2019

Abstract

Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroup. Using the dermatologist approved Fitzpatrick Skin Type classification system, we characterize the gender and skin type distribution of two facial analysis benchmarks, Labeled Faces in the Wild and Adience. We find that these datasets are overwhelmingly composed of lighter-skinned subjects (76.6% for LFW and 86.2% for Adience) and introduce a new facial analysis dataset which is balanced by gender and skin type. We evaluate 2 commercial gender classification systems using our dataset and show that darker-skinned females are the most misclassified group with error rates of up to 34.7%. The maximum error rate for lighter-skinned males is 0.6%. The substantial disparities in the accuracy of classifying darker females, lighter females, darker males, and lighter males in gender classification systems require urgent attention if commercial companies are to build genuinely fair, transparent and accountable facial analysis algorithms.

Buolamwini, J. & Debra, T. (2019) Gender shades: Intersectional accuracy disparities in commercial gender classification in Conference on fairness, accountability and transparency (pp. 75-81) FAccT.



資料查詢
創新設計營

博士
李佩珊
李佩珊
李佩珊

看見與歧視 -
ChatGPT談起

2023年11月13日 10:30-12:00
淡江大學資訊大樓講堂

講者A - 李佩珊



淡江時報