

成大心理系系友分享 資料科學家的日常

2017/12/09 游為翔 @ NCKU

Who am I



I'm HERE!

背景

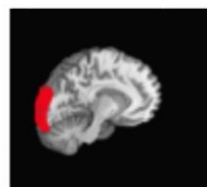
國立成功大學心理系 (101級)
- 知覺與功能性磁共振造影實驗室

國立台灣大學心理所
- Vision NeuronScience Lab

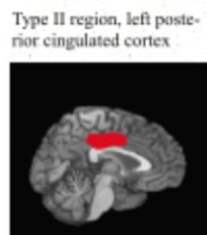
中央研究院資訊科學所
- 資料洞察實驗室, Data scientist

專長

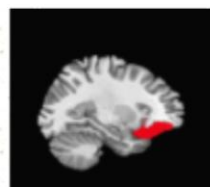
- 神經科學
- 機器學習與深度學習模型
- 消費者及使用者行為分析
- 深度學習於影像辨識之應用



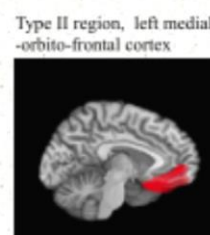
Type I region, left lateral-occipital cortex



Type II region, left posterior cingulate cortex



Type I region, left lateral-orbito-frontal cortex



Type II region, left medial-orbito-frontal cortex

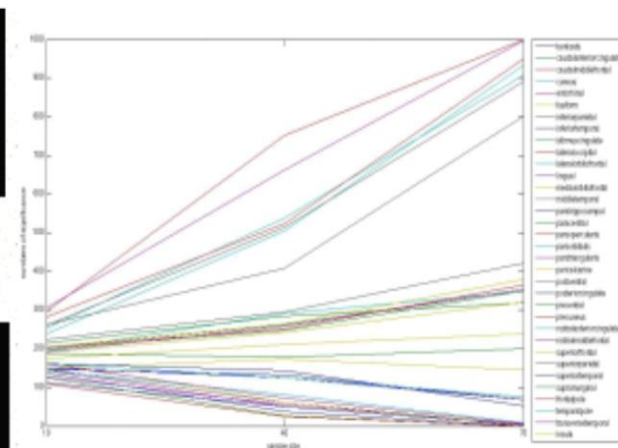
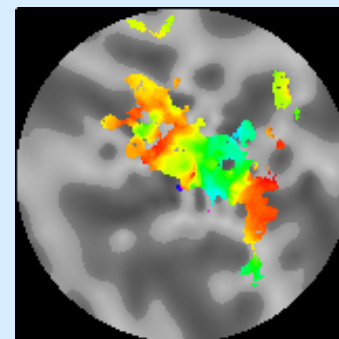
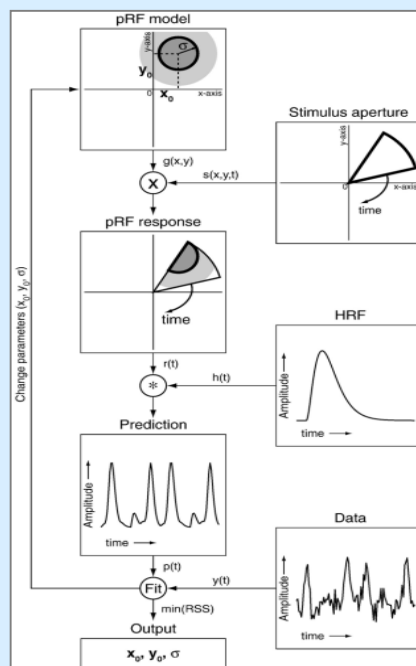
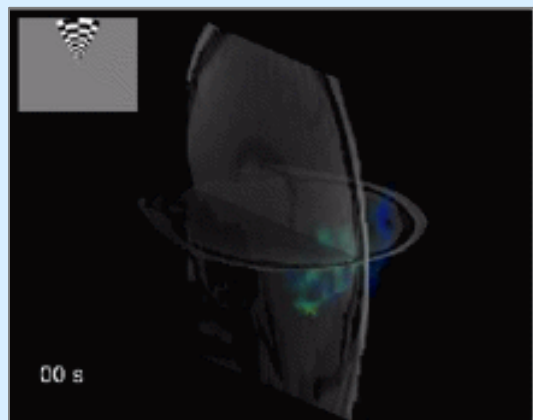
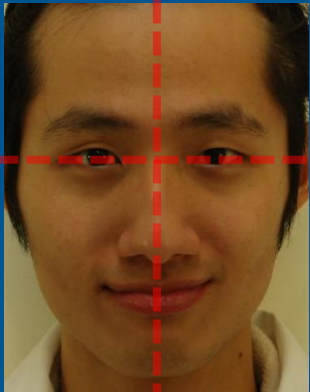


Fig. 1 ROIs result of numbers of significance in left hemisphere



1. Automatically face detection and alignment



2. Facebook Zombie Detector



為什麼進入資料科學



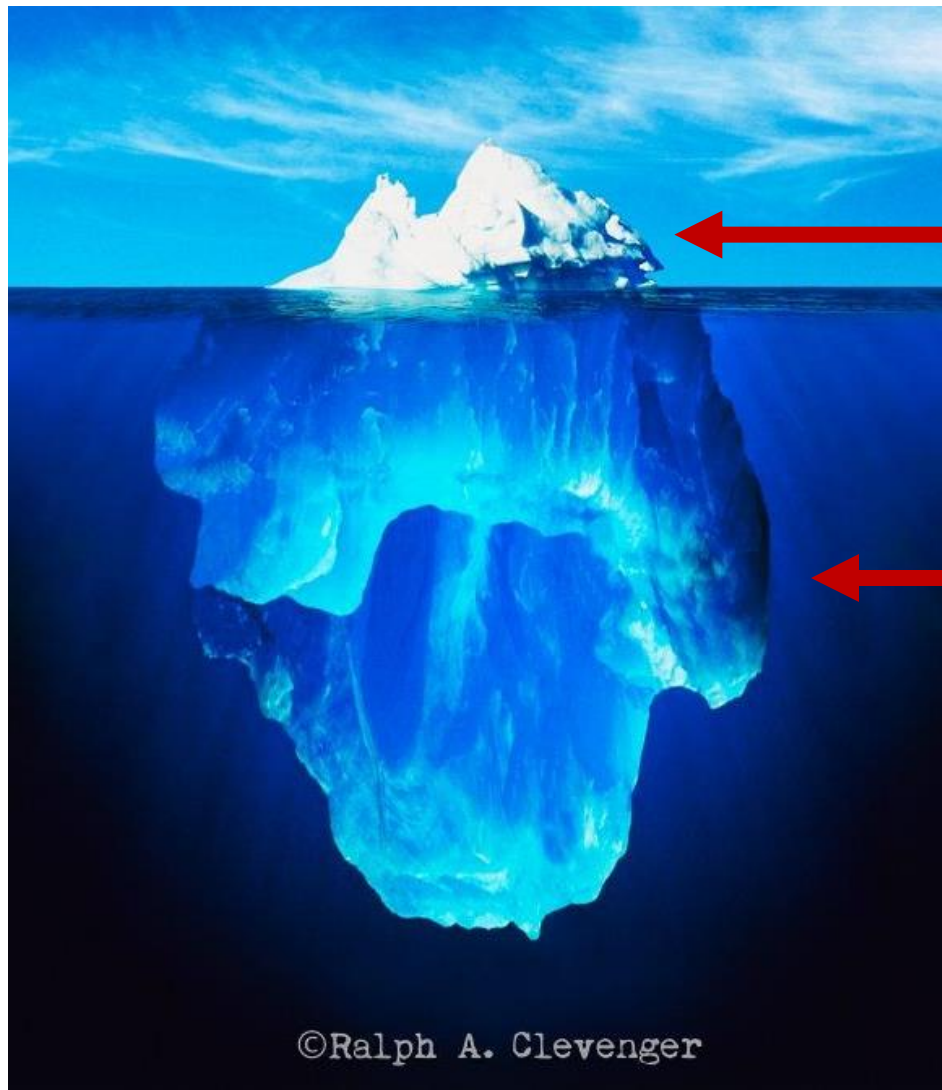
[When asked why he wanted to climb Mount Everest:] Because it's there.

(George Leigh Mallory)

izquotes.com

資料科學家平常都在做啥？

每日好時光



問客戶資料怎麼可以這麼髒

清理資料



資料收集



資料清理



資料分析



特徵值萃取



資料視覺化



建立模型



模型驗證



決策應用

清資料真的 hen 痛苦

那群探索未知的人：資料科學家

< 1 性感背後，資料科學家為什麼迷人？

2 「感覺資料都在嘲笑我了。」以博客來為例，看資料科學專案會遇到的困難

3 走進農田裡的資料科學家，吳君孝：資料是一門探索未知的學問 >

2

大數據

「感覺資料都在嘲笑我了。」以博客來為例，看資料科學專案會遇到的困難

by  曾毅 2017.06.07



圖片來源：蔡仁錫攝

86586^||^1^||^277^||^1^||^0^||^0^001064

合作專案



悠遊卡股份有限公司
EASYCARD CORPORATION



Sunfun



玉山銀行
E.SUN BANK



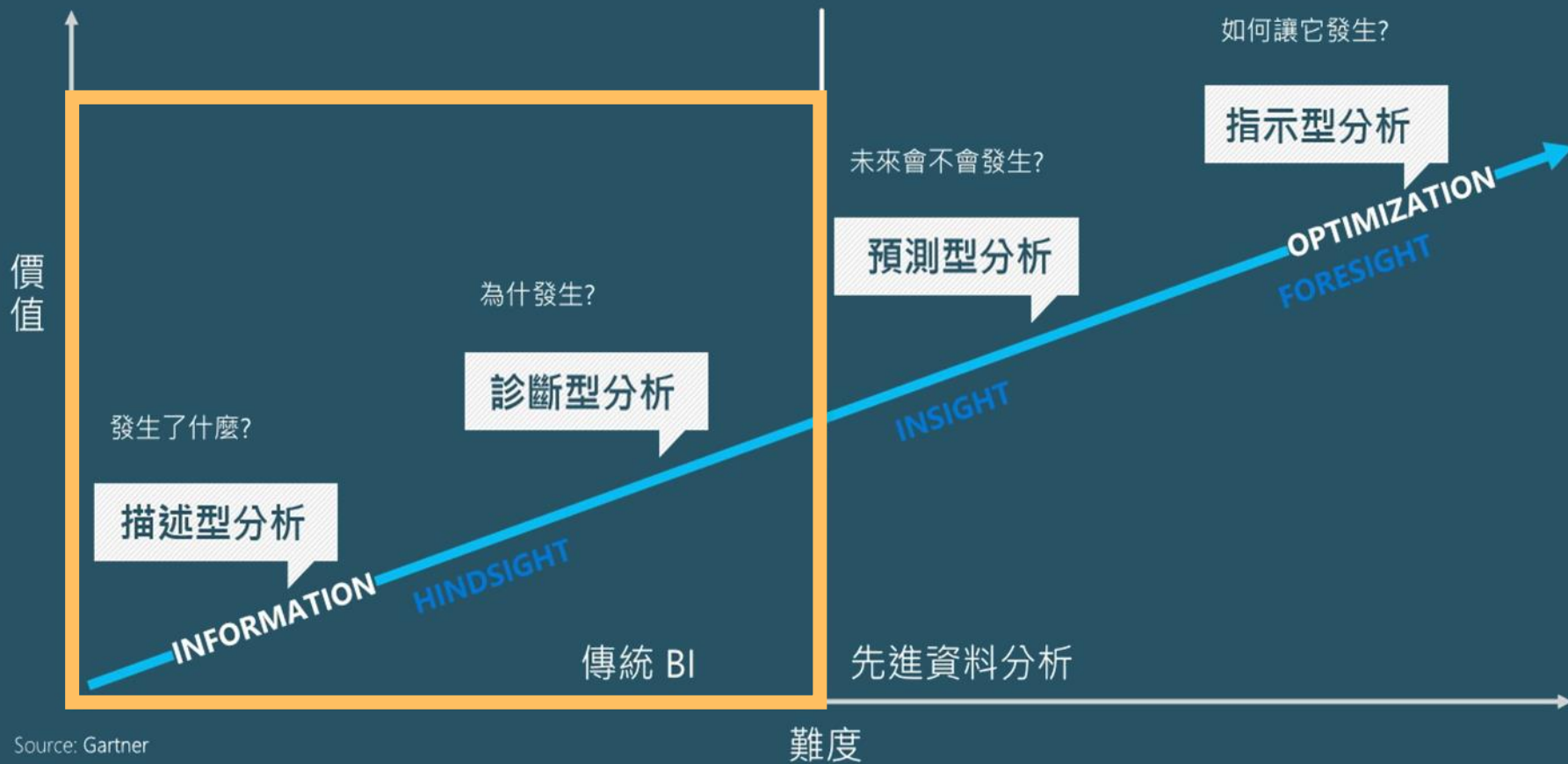
台塑企業
FORMOSA PLASTICS GROUP



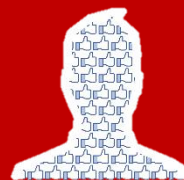
奇美電子
CHI MEI OPTOELECTRONICS

Some Example in 資料洞察實驗室





誰按的粉絲團?



You are what you like!

??

??

小茉莉-陳瑀希

Catworld小舖

Garena 《英雄聯盟 LOL》

EYESCREAM Inc.

波多野結衣Hatano Yui

LOVFEE

解婕翎

OB嚴選

豆花妹 蔡黃汝

grace gift

Nono_辜莞允

BEVY C.

張景嵐

Joyceshopstyle

FHM Taiwan 男人幫國際中文版

QUEEN FASHION SHOP

潮物blog - 街頭潮流男著

SweeSa水莎

張小筑 Ya Chu

Lulus

性別

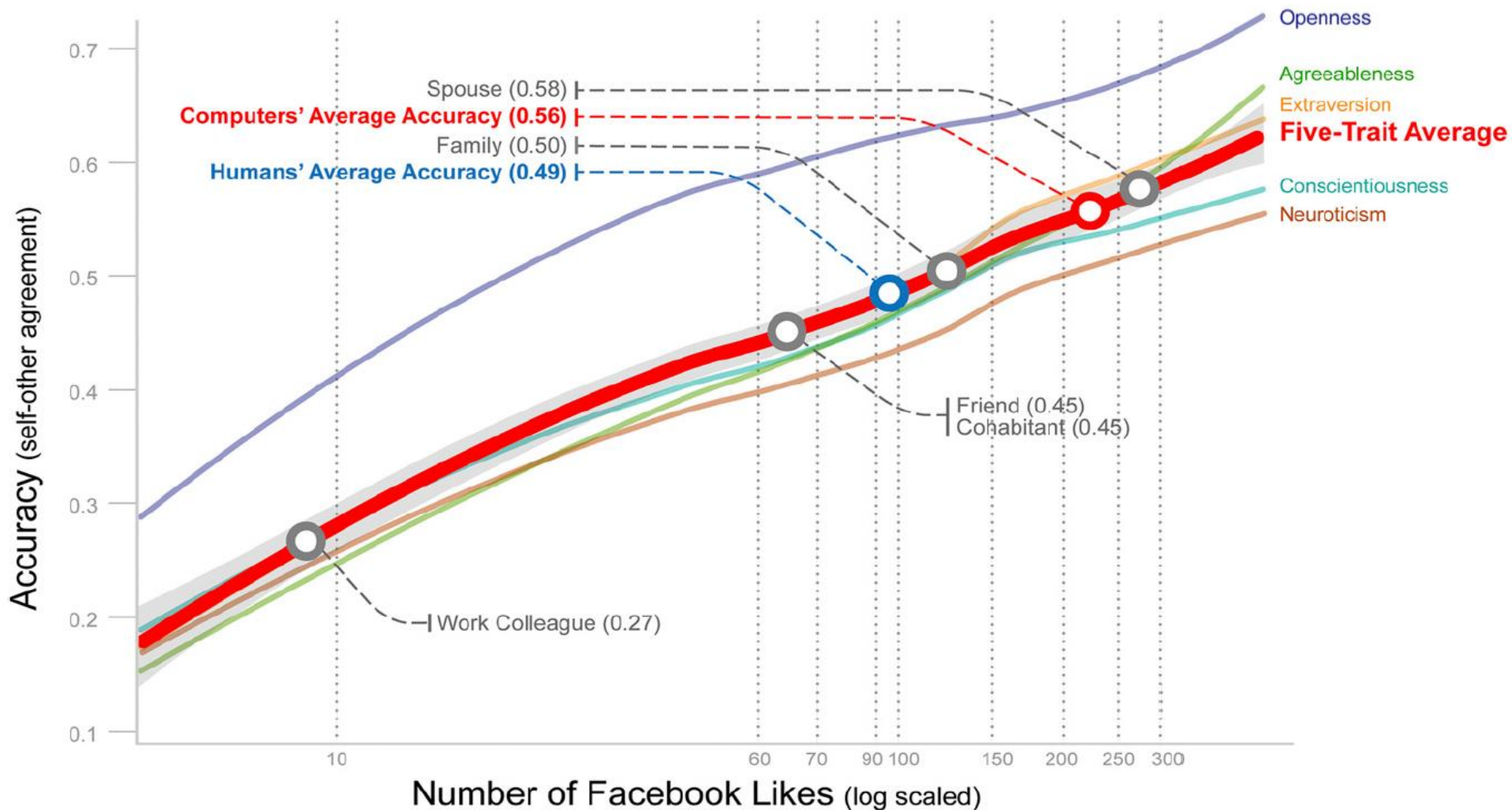
男	女
<u>小茉莉-陳瑀希</u>	<u>Catworld小舖</u>
<u>Garena 《英雄聯盟 LOL》</u>	<u>EYESCREAM Inc.</u>
<u>波多野結衣Hatano Yui</u>	<u>LOVFEE</u>
<u>解婕翎</u>	<u>OB嚴選</u>
<u>豆花妹 蔡黃汝</u>	<u>grace gift</u>
<u>Nono_辜莞允</u>	<u>BEVY C.</u>
<u>張景嵐</u>	<u>Joyceshopstyle</u>
<u>FHM Taiwan 男人幫國際中文版</u>	<u>QUEEN FASHION SHOP</u>
<u>潮物blog - 街頭潮流男著</u>	<u>SweeSa水莎</u>
<u>張小筑 Ya Chu</u>	<u>Lulus</u>

收入 (25 < age < 40)

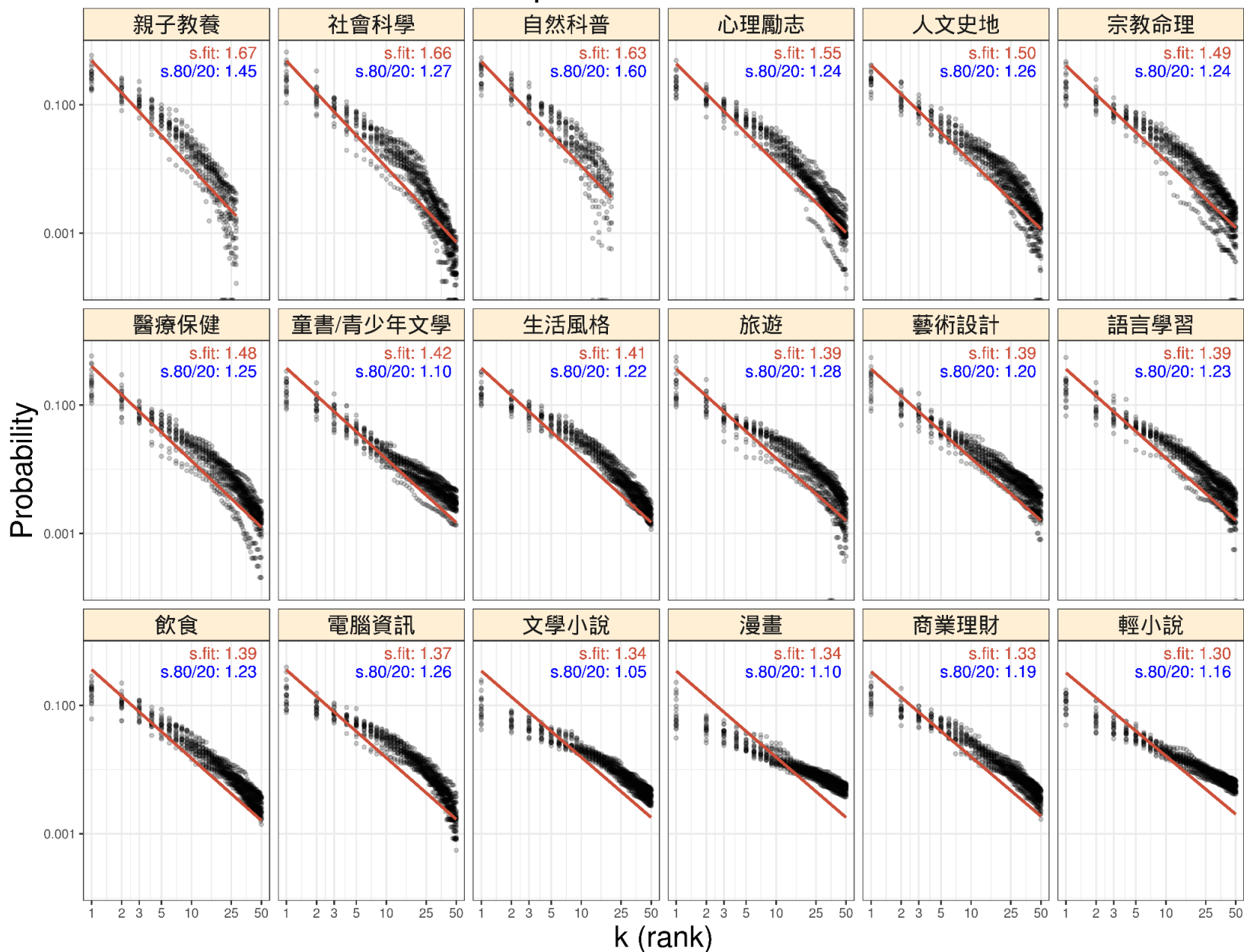
高	低
citiesocial	唯舞獨尊(臉書版)_首款社群平台音樂遊戲
商業周刊	Android 免費遊戲區
Mobile01	ENL 簡單新生活
財報狗	Tetris Battle
Technews 科技新報	學生愛打工
台灣廉價航空福利社	RC語音
Inside 硬塞的網路趨勢觀察	終極一班2
經理人月刊MANAGERtoday	終極x宿舍
Sway房市觀測站	遊戲大亂鬥
背包客棧	蔡芷紘

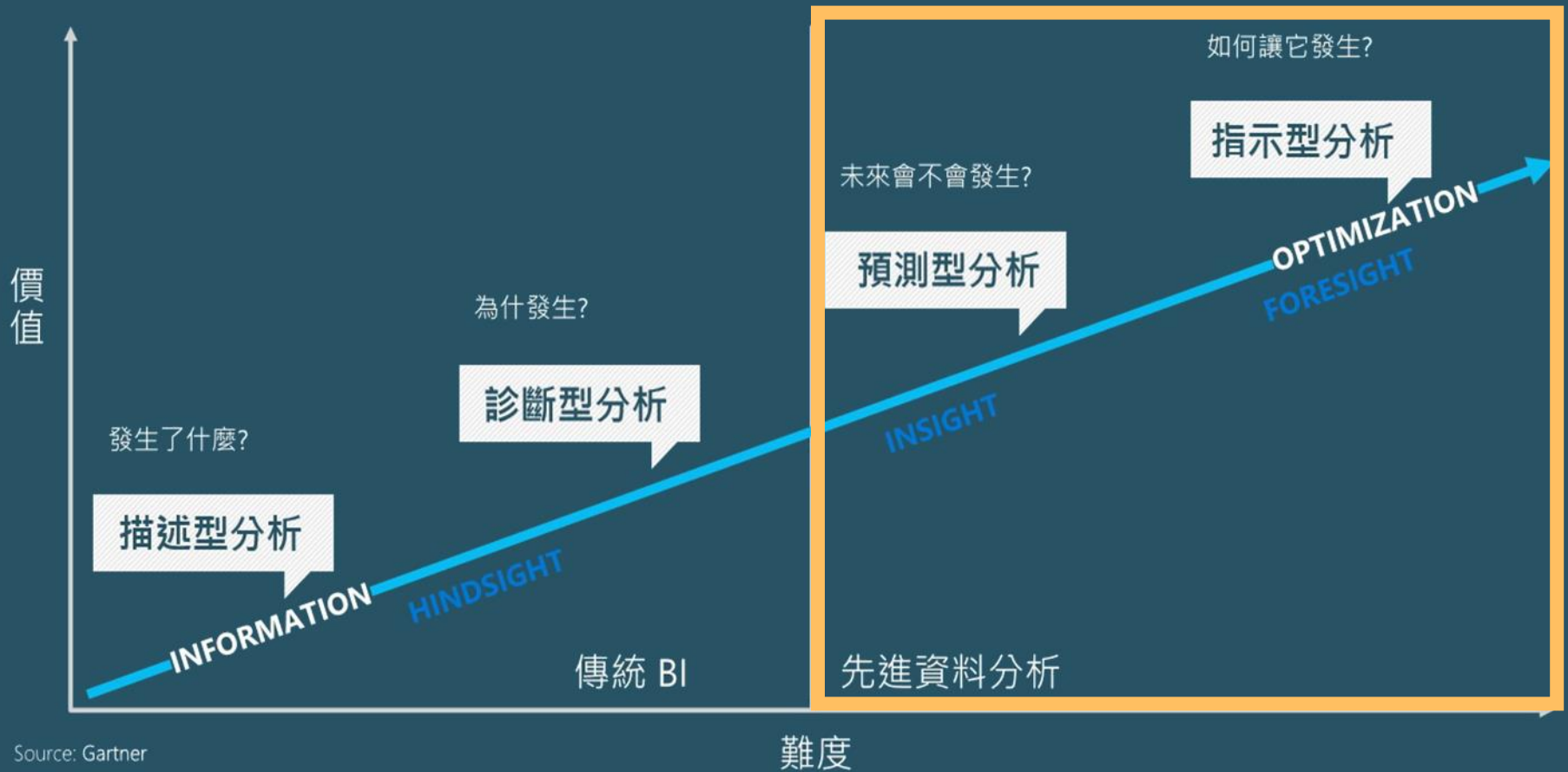
Computer vs. Humans

- 電腦用你按過讚的粉絲團來猜你的個性



The Zipf law in books market





實驗室轉型：轉向人工智慧應用與研究

Describes without errors



A person riding a motorcycle on a dirt road.

Describes with minor errors



Two dogs play in the grass.

Somewhat related to the image



A skateboarder does a trick on a ramp.



A group of young people playing a game of frisbee.



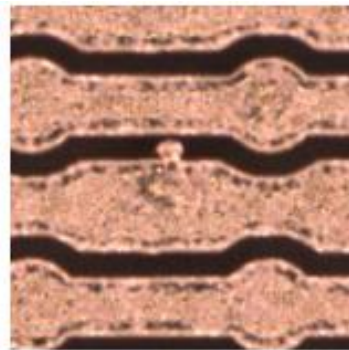
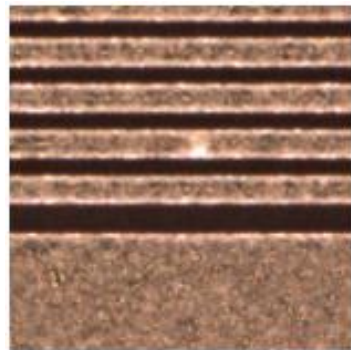
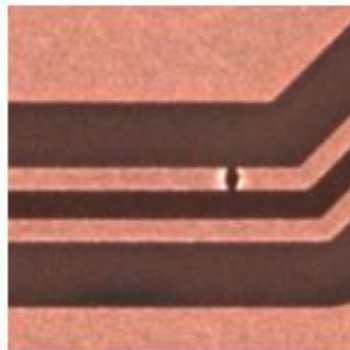
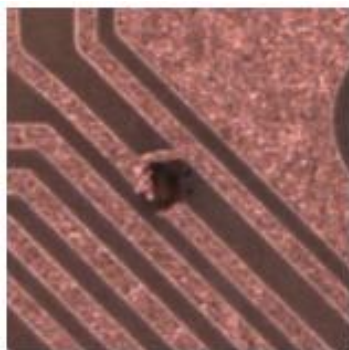
Two hockey players are fighting over the puck.



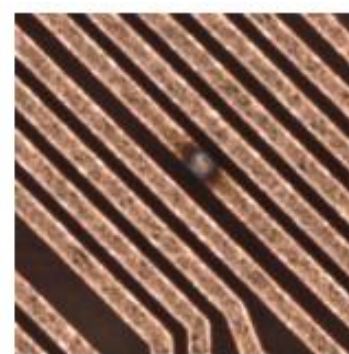
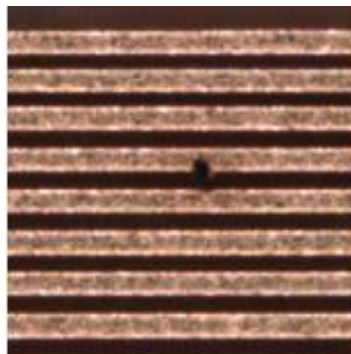
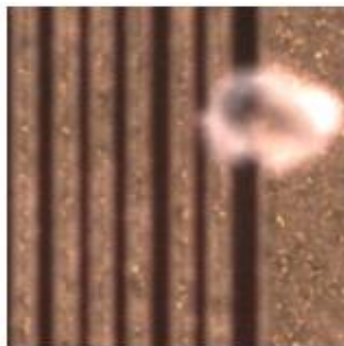
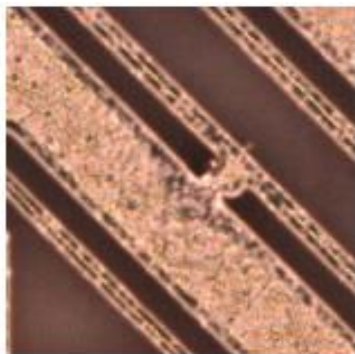
A little girl in a pink hat is blowing bubbles.

工業影像 - 缺陷分類

有特定瑕疵



無特定瑕疵



目標物追蹤：追蹤埋葬蟲移動路徑

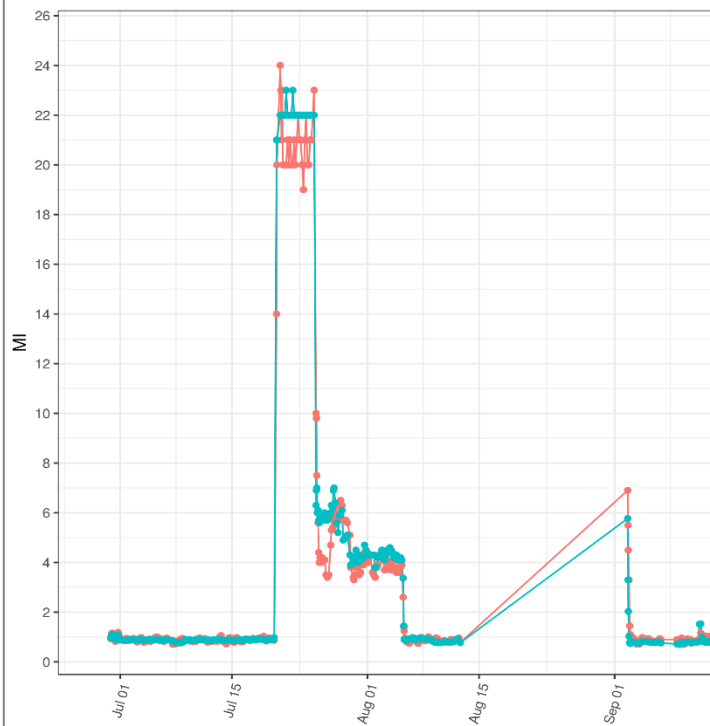


參數調控：化學產品品質控管－預測與控制

MI 預測結果

時段: 2016-06-29 19:30:00 ~ 2016-11-09 11:00:00

— 實際值 — 預測值



預測與實際值之相關係數可達
0.94 (相較過去傳統方法為 0.78)

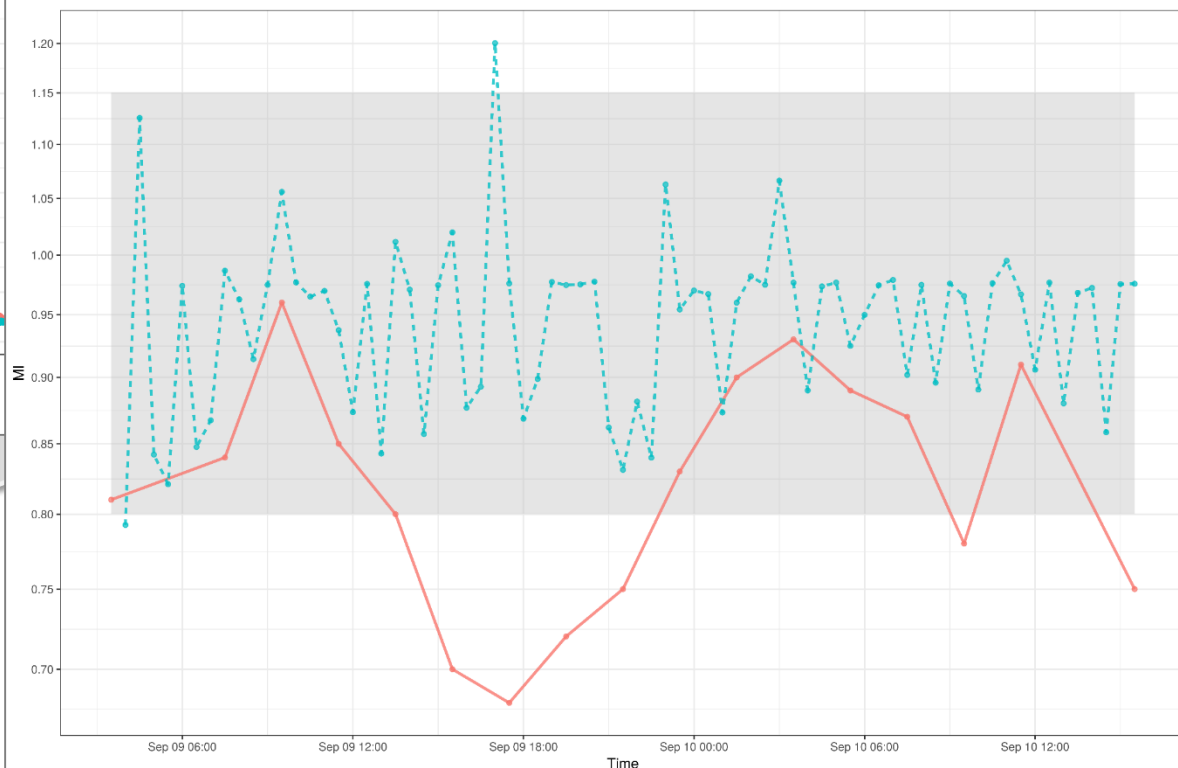
測量標的：熔融指數 (MI)

轉品階段 MI V234 之變化

轉品目標: 8010

時段: 2016-09-09 03:30:00 ~ 2016-09-10 15:30:00

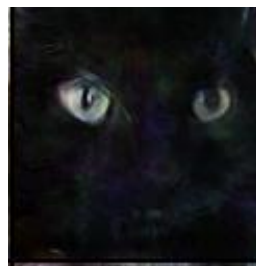
— 人為操作之檢測結果 — 智能控制之預測結果



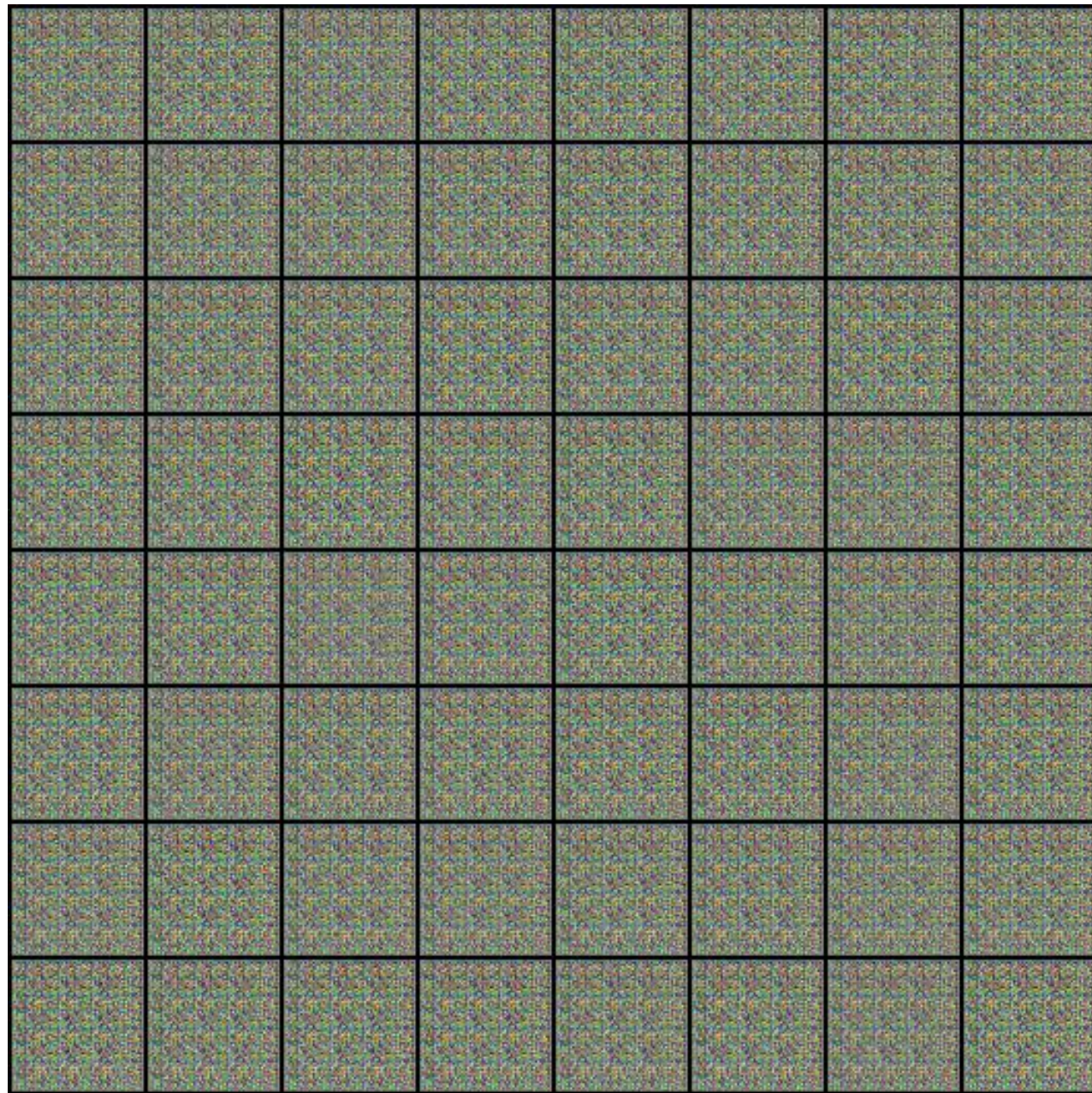
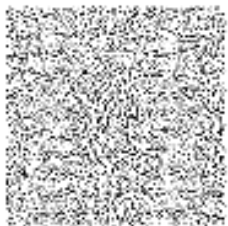
MI 在目標範圍時間比例
人為操作: 61.1%
智能控制: 98.6%

其他特殊應用

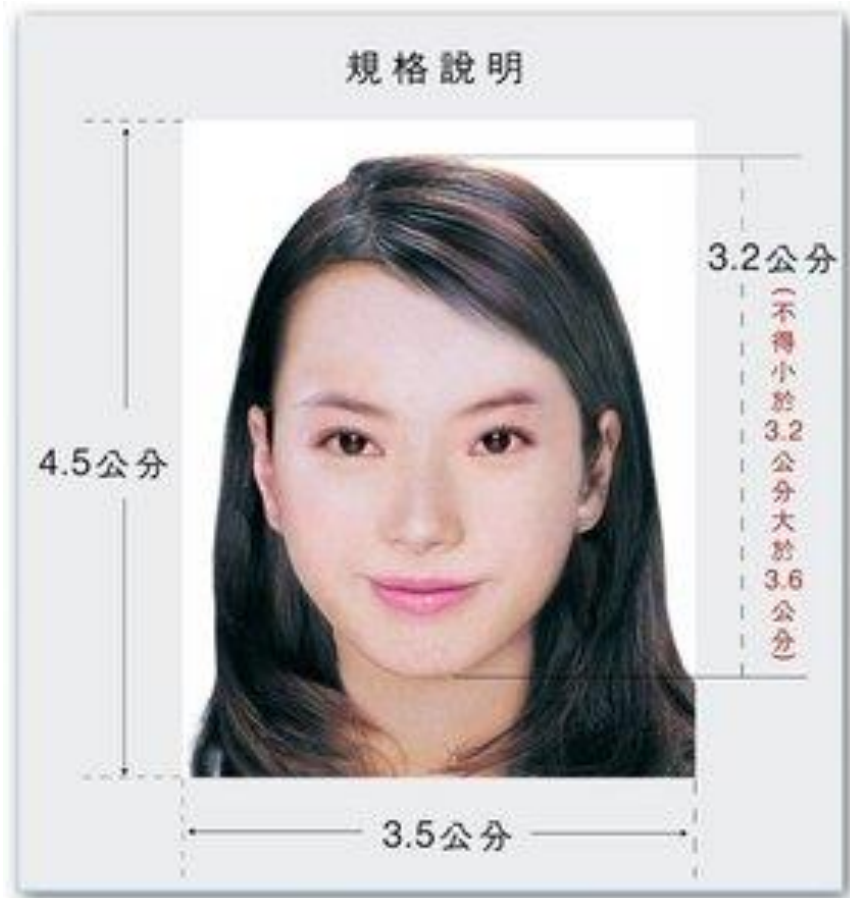
假圖生成研究：哪隻貓是假的



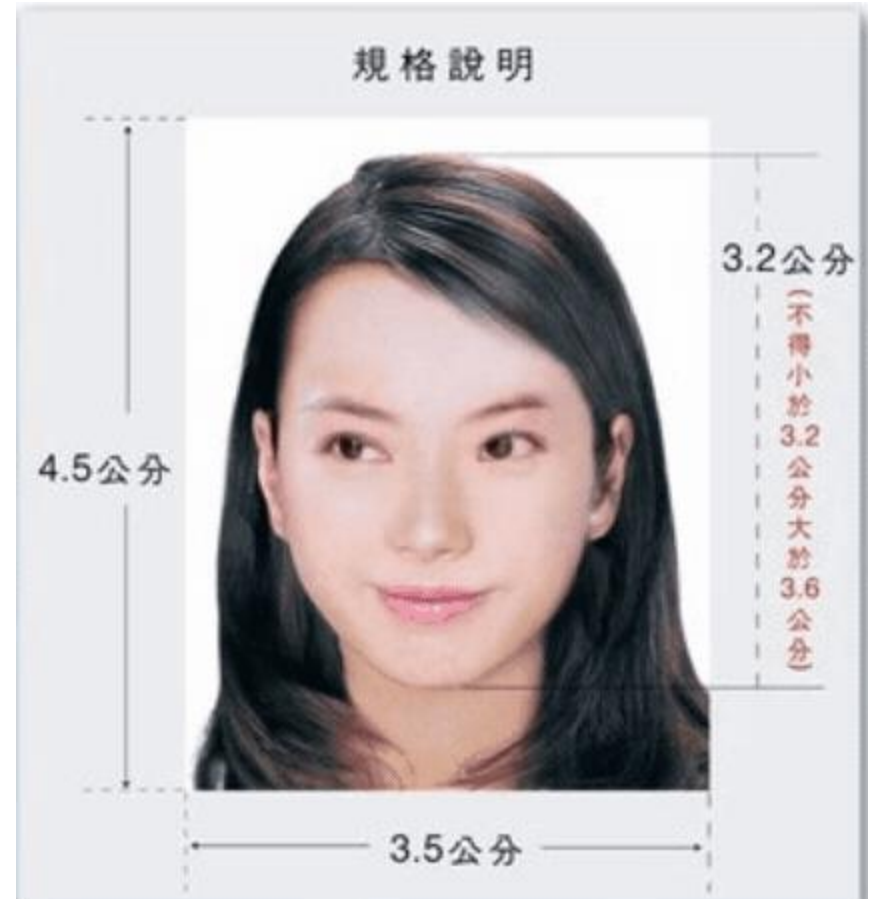
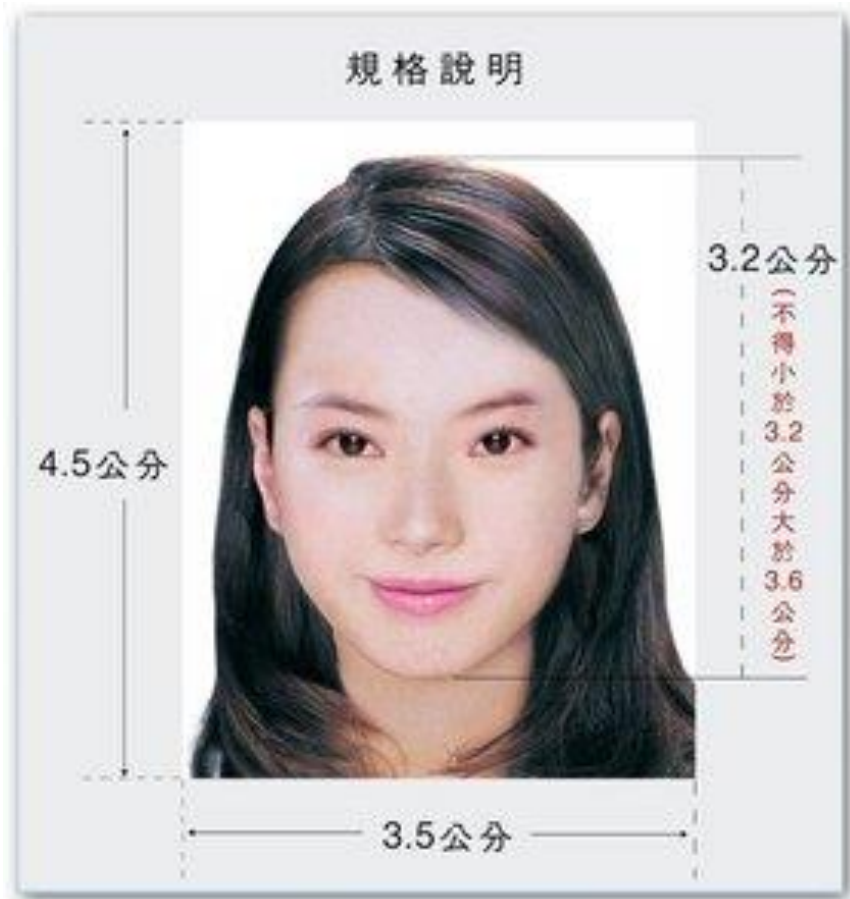
Random
noise

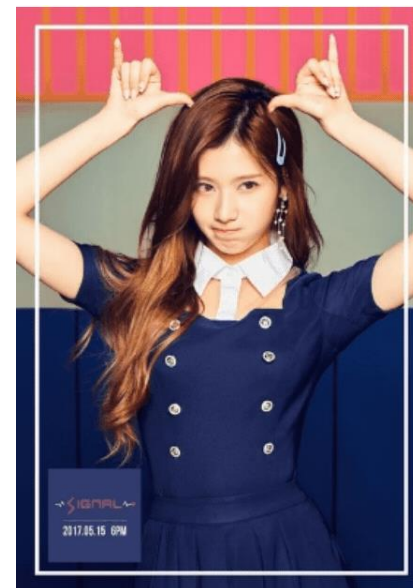
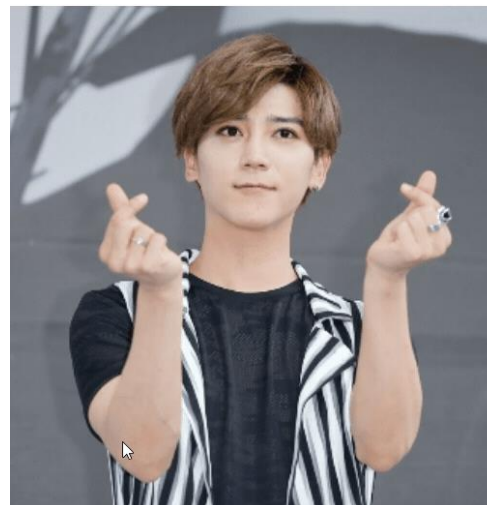


影像生成研究



影像生成研究





從心理系該如何跨入資料科學

從心理系該如何跨入資料科學

- 學會寫程式
 - 什麼程式都沒關係, 你要能用程式語言實現你的想法
- 多接觸各種領域的課程
 - 生科、經濟、物理化學、資訊, etc.
- 從做中學：去 Kaggle 上玩玩

kaggle

Code

codecademy



DATAQUEST



DataCamp

MOOC

coursera



UDACITY



從心理系該如何跨入資料科學

- 別忘記你 (在心理學) 的專業

- 所以台灣現在的狀況是怎麼回事？

台灣近年來創業聲浪潮起，意味著創業相關之法律、稅務等議題都逐漸受到重視。台灣過去十幾年來，各類培育計畫紛紛成立，創業二字在台灣並不罕見。

然而台灣過去幾年在科技創業上的努力，成功案例仍多侷限於國外成功模式的翻版。過去數年台灣本土成功案例西進中國多不敵當地競爭對手之資本規模，近年來才有較多轉戰東南亞和歐美國家的趨勢。

這種現象告訴我們，**台灣的科技新創公司原創性低。而上溯源頭，恐怕要徹底檢討產學合作機制、研發投資規模、科技基礎建設以及科技跨領域 (Interdisciplinary) 合作**。打個比方，心理學和分析哲學不強的國家，使用者經驗 (User Experience)、人工智慧、機器人學等領域很難強起來，在這些領域強勢的國家背後，都鉅額投資多學術領域與相關產業的合作並進，不是光投資工程就可以突破。

中文 英文 日文 偵測語言

↔

英文 中文(繁體) 中文(簡體)

翻譯

Do yuo fnid tihs smilpe to raed?
Beceuae of the phaonmneal
pweor of the hmuan mnid, msot
plepoe do.

100/5000

你是不是該嘲笑這個人？ 這個人的
mnid的哼哼唧唧唧唧唧唧唧唧唧唧
的。

Nǐ shì bùshì gāi cháoxiào zhègè rén? Zhègè rén de mnid de hēng
hēng jī jījījījī jī jī de.

你是不是要查： Do yuo fnid tihs smilpe to *read*?
Because of the phaonmneal pweor of the hmuan mnid,
most people do.

Do yuo fnid tihs
smilpe to raed?
Bceuae of the
phaonmneal pweor
of the hmuan mnid,
msot plepoe do.

Jumbled letter
Warrington, 1980

The diagram illustrates the experimental setup for Hubel and Wiesel's study. A rabbit's head is shown in profile, facing left. A recording electrode is inserted into the visual area of the brain. A line labeled 'from brain' connects the electrode to a box labeled 'Stimulus'. The box has a screen displaying a vertical bar, representing the visual stimulus presented to the rabbit.

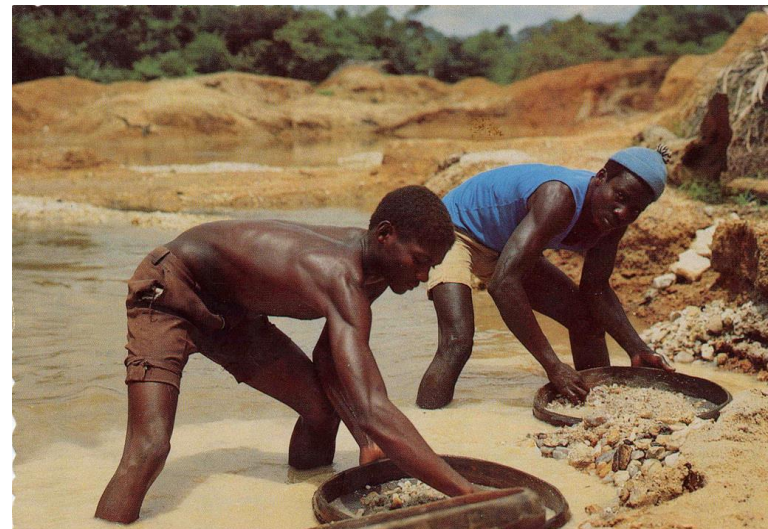
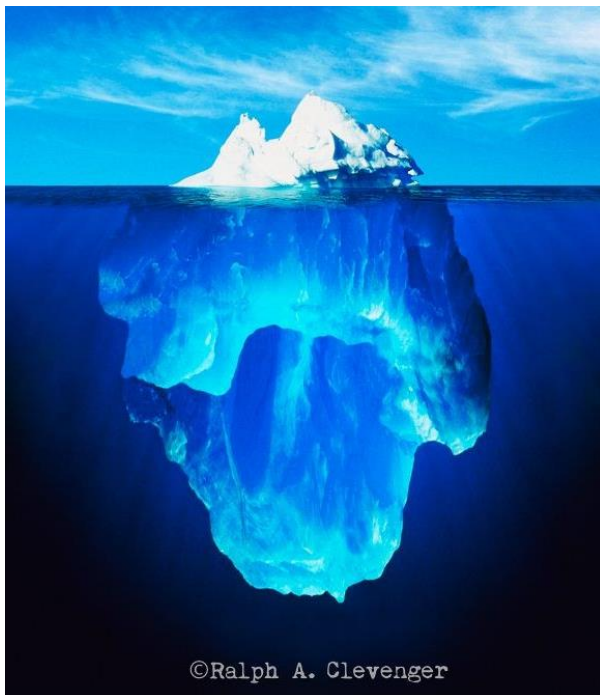
Receptive fields
Hubel & Wiesel, 1959

從心理系該如何跨入資料科學

- 培養溝通能力與觀察力
- 早點了解自己，清楚你的優勢
 - 去實習 (或去跑實驗室)，趁你還有機會反悔的時候反悔



資料分析中 **90%** 的時間是很枯燥乏味的，但當成功
找到有意義的結果時，會覺得努力都是值得的



整個過程中最難的地方在於
當你不知道這件事情是否可行時
你要能證明它可行

胡郁 <科大訊飛執行總裁>

Thank You!