



人工智慧與自然語言： 語文和工程專業之跨域生涯

- 講者：[江豪文博士](#)
- 日期：2024-12-09

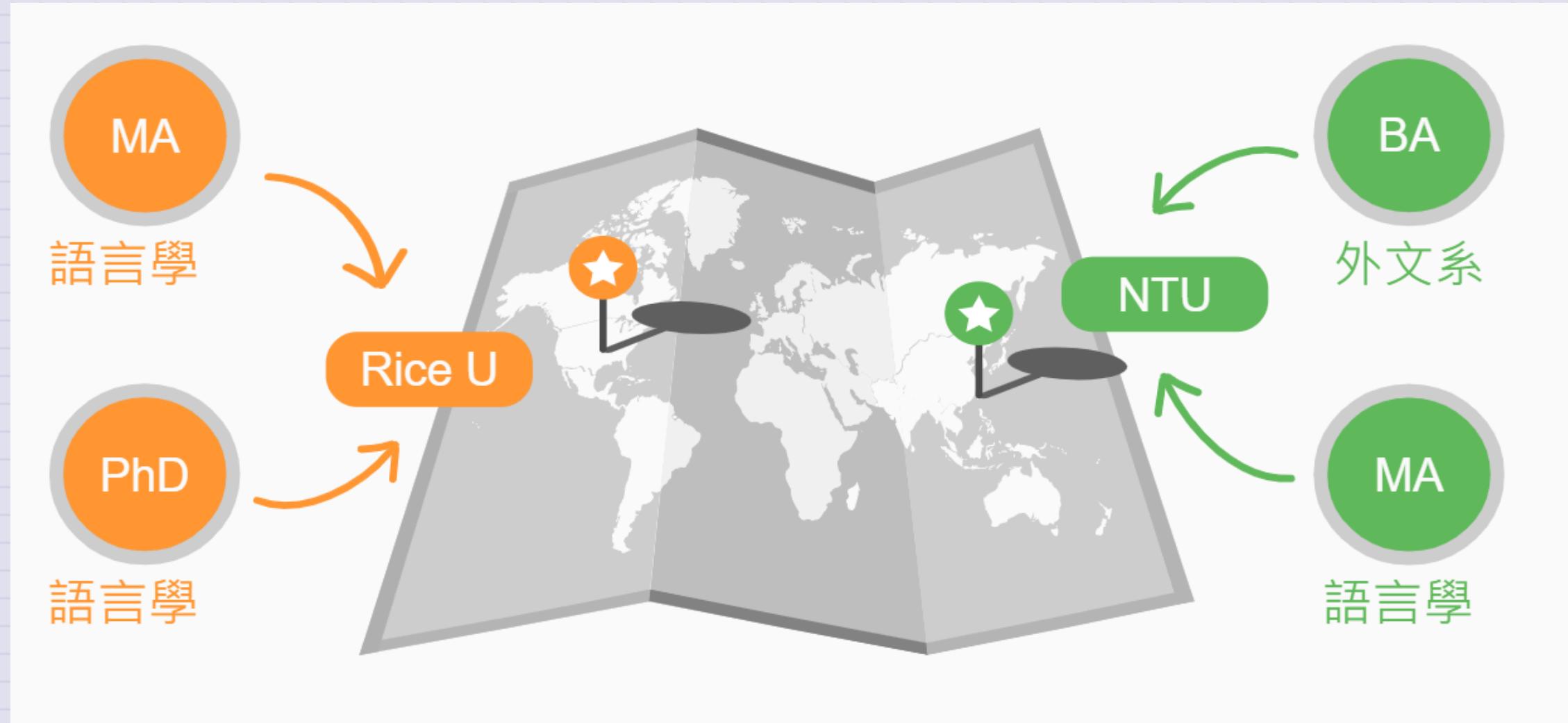
大綱

- 教育背景
- 工作經歷
- 轉職歷程
- 問答時間



教育背景





English Major

外文系
做什麼

語文和工程專業之跨域生涯



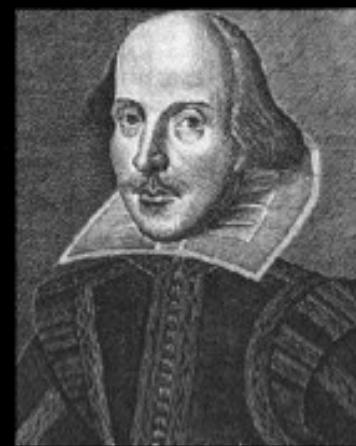
What my friends think I do



What my parents think I do



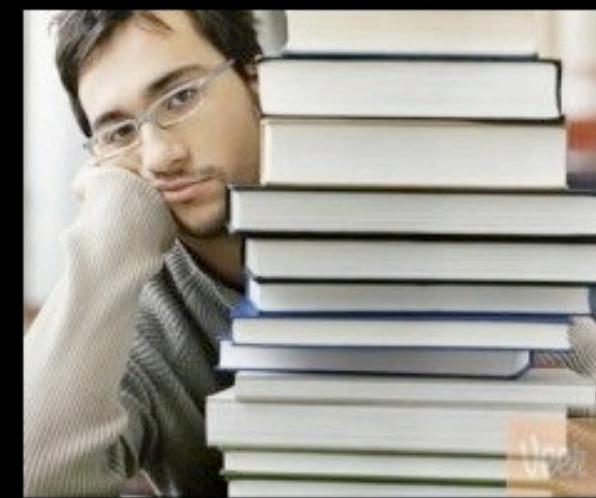
What my professors think I do



What society thinks I do



What I think I do



What I actually do

Linguistics Student

語言學
學什麼

語文和工程專業之跨域生涯



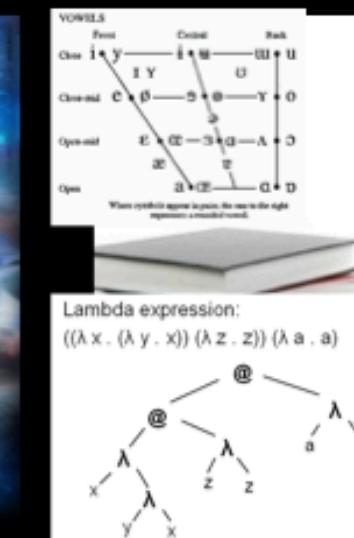
What society thinks I do.



What I think I do.



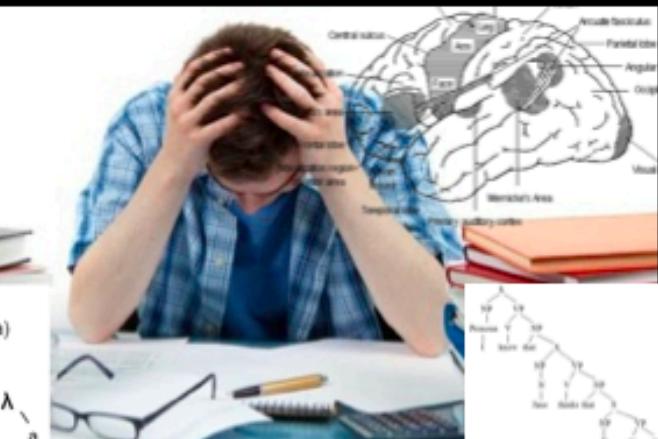
What my friends think I do.



What I really do!



What my parents think I do.

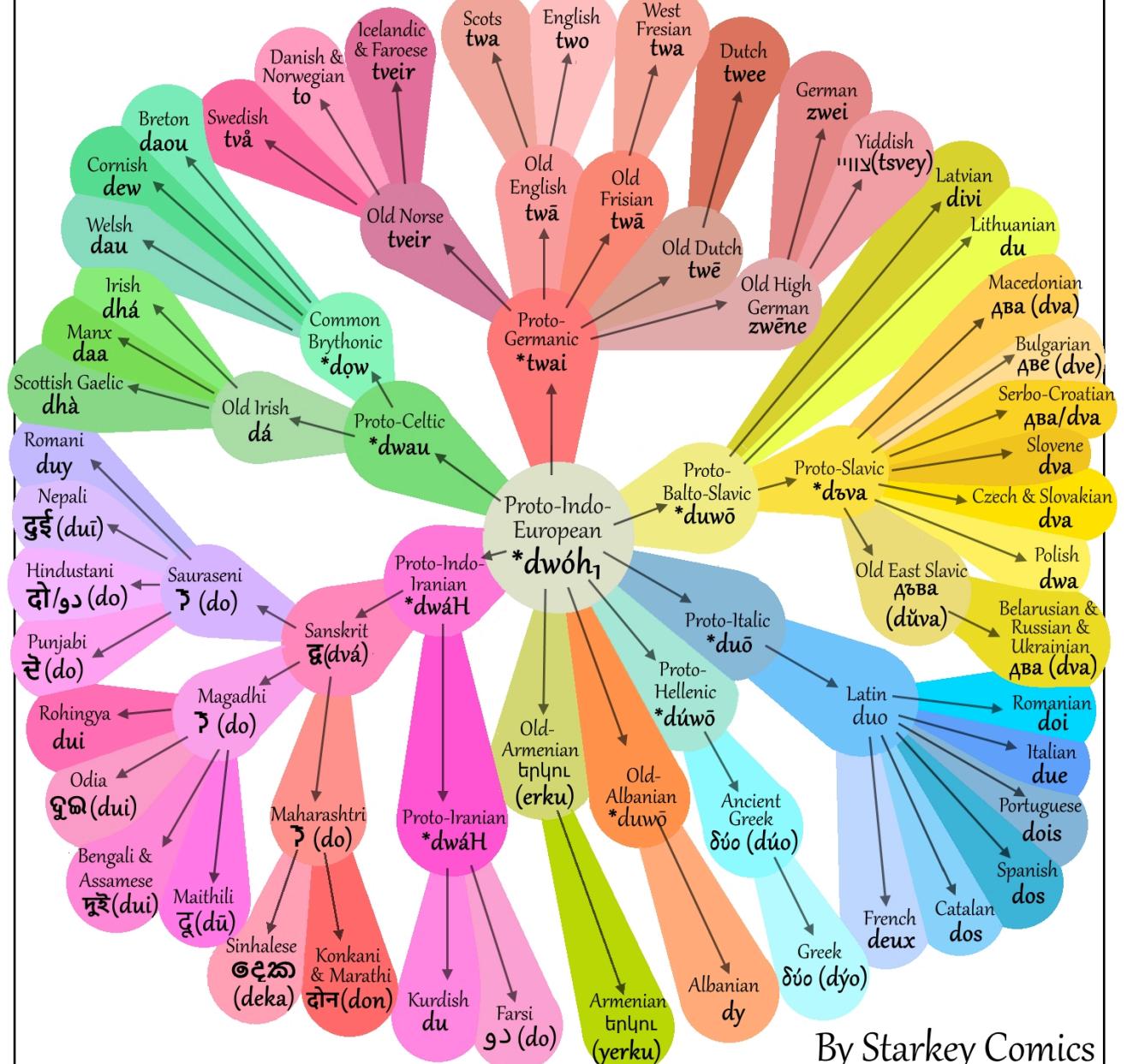


原來語言學 不是學語言

我以為我會 研究的語言

語文和工程專業之跨域生涯

Indo-European Words for Two



但我實際上 研究的語言

語文和工程專業之跨域生涯





台灣南島語

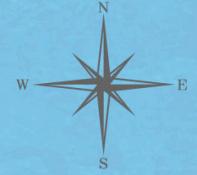
Formosan Languages

語文和工程專業之跨域生涯

臺灣原住民族 16族分布圖

臺灣原住民族文化園區

Namasan Kakivaualjan tua Kakudakudan nua Kacalision i Taiwan

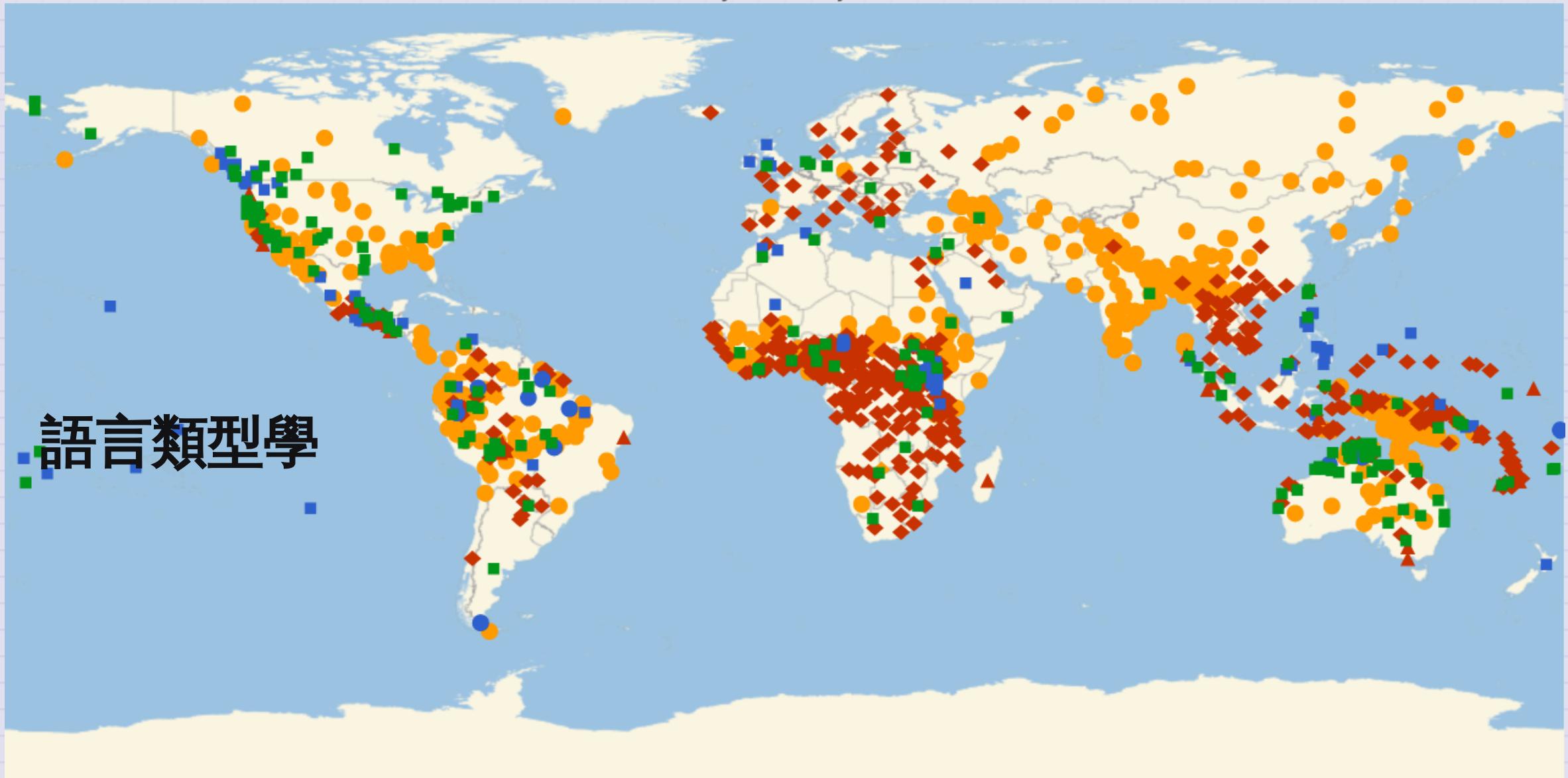


學位論文

- 碩士: 噶瑪
蘭語空間認
知之研究
- 博士: 台灣
南島語的名
物化與領屬
結構



Order of Subject, Object and Verb



- 1 SOV
- 2 SVO
- 3 VSO
- 4 VOS
- 5 OVS
- 6 OSV
- 7 No dominant order

- 名物化論文 1/2

Jiang, Haowen. 2021a.

Argument nominalization in
Formosan languages: A
functional-typological
approach. Osaka: Osaka
University Press.

[Amazon >>](#)

Nominalization Theory
and
Linguistic Analysis

Sung-Yeo Chung
Masayoshi Shibatani

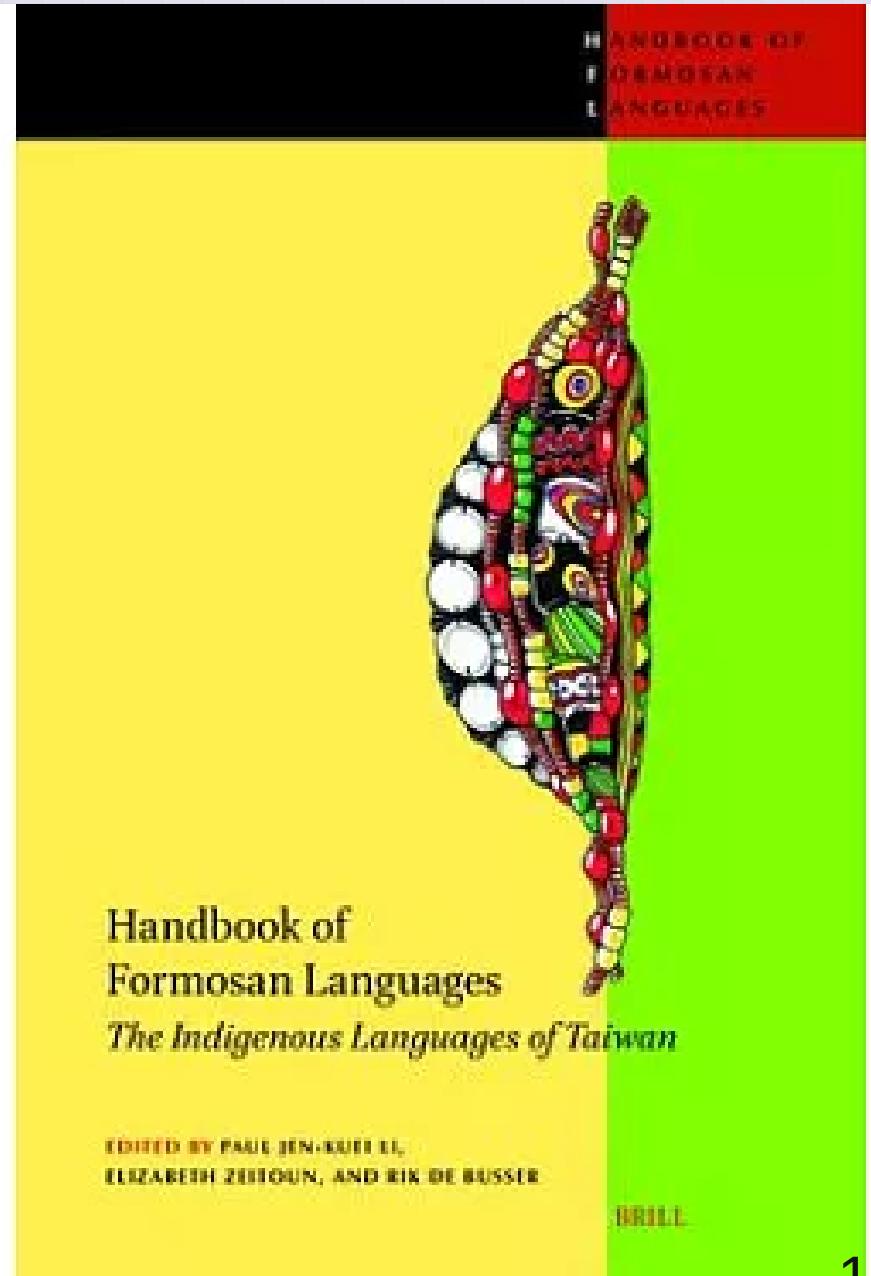
体言化理論と
言語分析

鄭聖汝
柴谷方良
編

- 名物化論文 2/2

Jiang, Haowen. 2023b. 16.
Nominalization in
Formosan Languages.
Leiden: Brill

[Brill >>](#)



- 布農語線上詞典 >>

The screenshot displays the homepage of the Indigenous Languages Online Dictionary. On the left, a vertical green sidebar features the title "原住民族語言線上辭典" and a stylized yellow leaf logo. The main content area shows circular icons for various languages, each accompanied by its name in Chinese. The languages listed are: 阿美語, 泰雅語, 排灣語, 布農語, 卑南語, 魁凱語, 鄒語, 賽夏語, 雅美語, 邵語, 噶瑪蘭語, 太魯閣語, 撒奇萊雅語, 賽德克語, 拉阿魯哇語, and 卡那卡那富語. To the right, there is a navigation bar with links to "網站導覽", "關於本站", and "使用手冊下載". A "最新消息" section contains a notice about a language dictionary recruitment drive. Below it, a "每日一句" section shows a phrase in噶瑪蘭語 with its English translation. A "跨族語檢索" search bar is also present.

原住民族語言線上辭典

阿美語 泰雅語 排灣語 布農語

卑南語 魁凱語 鄒語 賽夏語

雅美語 邵語 噶瑪蘭語 太魯閣語

撒奇萊雅語 賽德克語 拉阿魯哇語 卡那卡那富語

... 網站導覽 · 關於本站 · 使用手冊下載

最新消息

【徵的就是你(妳)！】族語線上辭典招募
約聘人員 😊

原住民族語言研究發展基金會「111-112年族語線上
辭典維護及系統功能.....

more

噶瑪蘭語

puniRan ku zanum na bekisi zau.
我把這個水桶加滿水了。

跨族語檢索：

族語選擇：不拘

請輸入中文或族語

起學族語

- 標記式語言 XML >>

```
<?xml version="1.0" encoding="UTF-8"?>
<interlinearGloss>
  <sentence>Я тебя люблю.</sentence>
  <gloss>
    <word>Я</word>
    <gloss>I</gloss>
    <translation>I</translation>
  </gloss>
  <gloss>
    <word>тебя</word>
    <gloss>you</gloss>
    <translation>you</translation>
  </gloss>
  <gloss>
    <word>люблю</word>
    <gloss>love</gloss>
    <translation>love</translation>
  </gloss>
  <freeTranslation>I love you.</freeTranslation>
</interlinearGloss>
```

 檢索

詞項列表

範疇分類

逐詞註解

首頁 / 布農語 /  檢索

Mazima

詞頻： (1118)

收錄來源：九階教材

[來源 >>](#)**解釋1** 喜歡

- Mazima saikin saicia. 

我喜歡牠。

-
- Mazima saikin maun mas haisingdulaz. 

我喜歡吃糯米蒸筒飯。

• 通過布農語認證考試
台灣立報 >>

語文和工程專業之跨域生涯

10月拚認證 找回母語記憶

記者：呂淑姮台北報導

地點：全臺 全部

「達悟」在達悟族語裡面，代表「人」的意思，指的是在蘭嶼的達悟族人。手上掛著象徵「船之眼」串珠的達悟族人廖清福說，他從15歲離開蘭嶼漁人部落，到台東念書，之後就很少再講族語。

今年53歲的廖清福，高中開始到台東求學，到外地之後沒有機會和環境使用自己的達悟母語，雖然小時候會講，但是當時碰到禁說方言政策，讓族語的回憶漸漸有些模糊。

近幾年來原住民族語復振，廖清福每到飛魚季，一定回到蘭嶼老家，向80多歲的母親認真學習母語，要把達悟族語再找回來，後來成為97年達悟族語認證考試榜首。現在，他也將族語傳承給自己的孩子：「希望他們知道自己的來處。」

行政院原民會18日舉辦「由我們發聲！母語是我們的身分證」記者會。副主委林江義宣布，99年度原住民族語言能力認證考試將於10月30日舉辦，報名日期自即日起到9月17日。

林江義說，今年是第8屆母語認證考試，和前7次不同的地方在於獎勵辦法。只要年滿70歲或者未滿15歲通過認證者、非原住民籍通過認證者、成績滿分、各族別榜首、6人以上團體報名，都可以獲得1千元獎金與表揚。

族語考試不限定年齡與族別，漢人也可以參加。一位曾在台大語言所擔任研究助理的學生江豪文，通過布農族語認證，他在台大時參與的研究和布農族語有關，覺得布農語很好聽，所以特別去學習，也通過了考試。

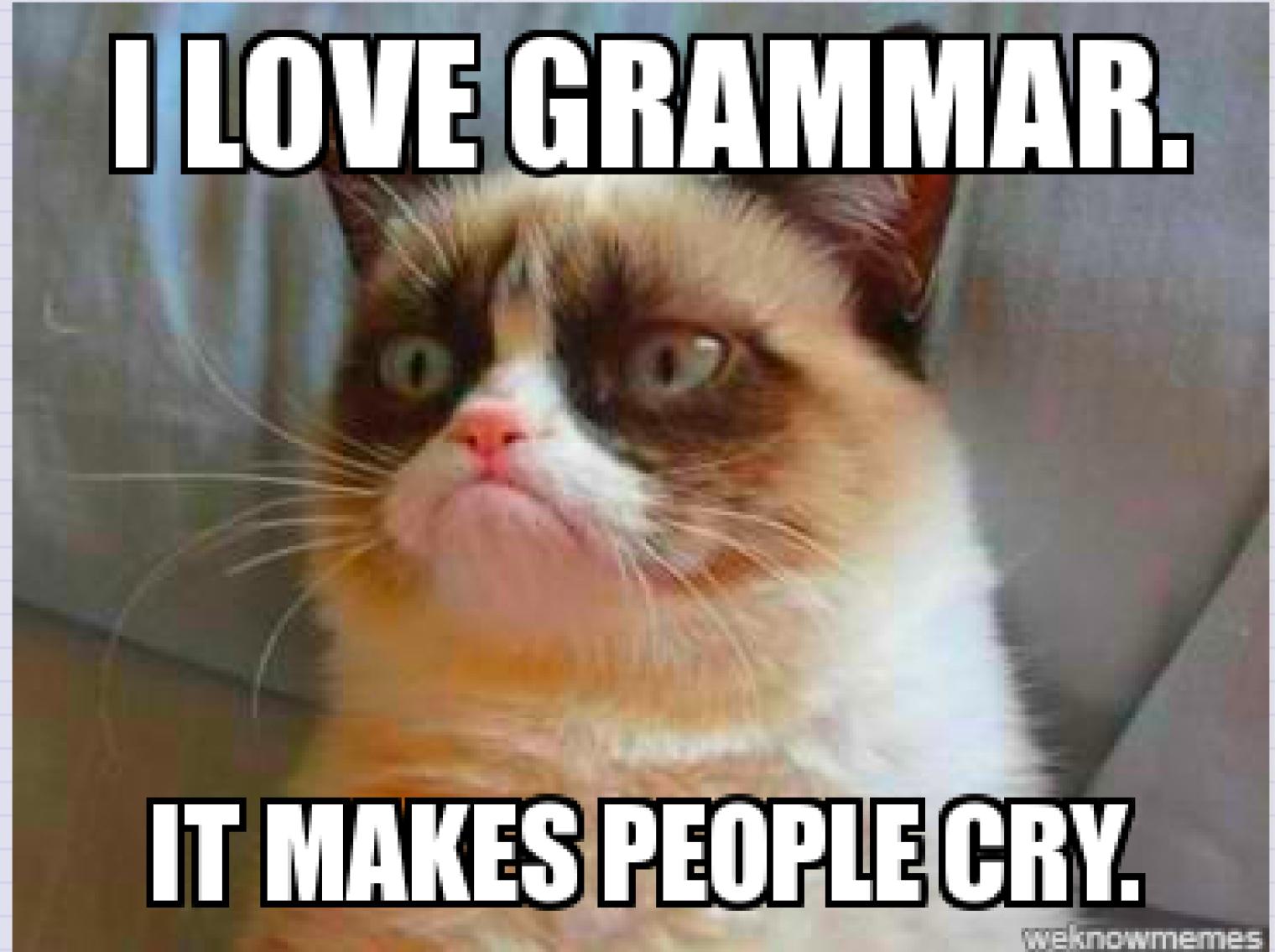
工作經歷





英文講師

臺北科技大學
應用英文系



博雅博士後 研究員

北京大學
中文系

語文和工程專業之跨域生涯

The screenshot shows the homepage of the Peking University Journals Platform. The header includes the university's logo and name in Chinese and English, along with navigation links for Home, Journals, Articles, and About, and options for Login/Registration and language selection. A large banner image features autumn foliage and traditional Chinese architecture. The main content area displays an article titled 'Essays on linguistics' with a sub-section 'Person Systems in Puyuma, an Austronesian Language of Taiwan'. The article details include the author (江豪文), affiliation (北京大学中国语言学研究中心/北京大学中文系 100871), and citation information. To the right, there are download links for the PDF (1712 KB), reading statistics (285 reads), and related articles sections.

北京大学 期刊网
Peking University Journals Platform

Home Journals Articles About
Login | Registration | 中文(简体)

Essays on linguistics

Title Please enter search term Search in journals Advanced

Journal homepage Introduction Instructions for Authors Back Issues

Essays on linguistics > 2023 , (04) : 20-40.

卑南语的人称标记

Person Systems in Puyuma, an Austronesian Language of Taiwan

Author(s) : 江豪文¹

Affiliation(s) : 1 北京大学中国语言学研究中心/北京大学中文系 100871

Cite this article

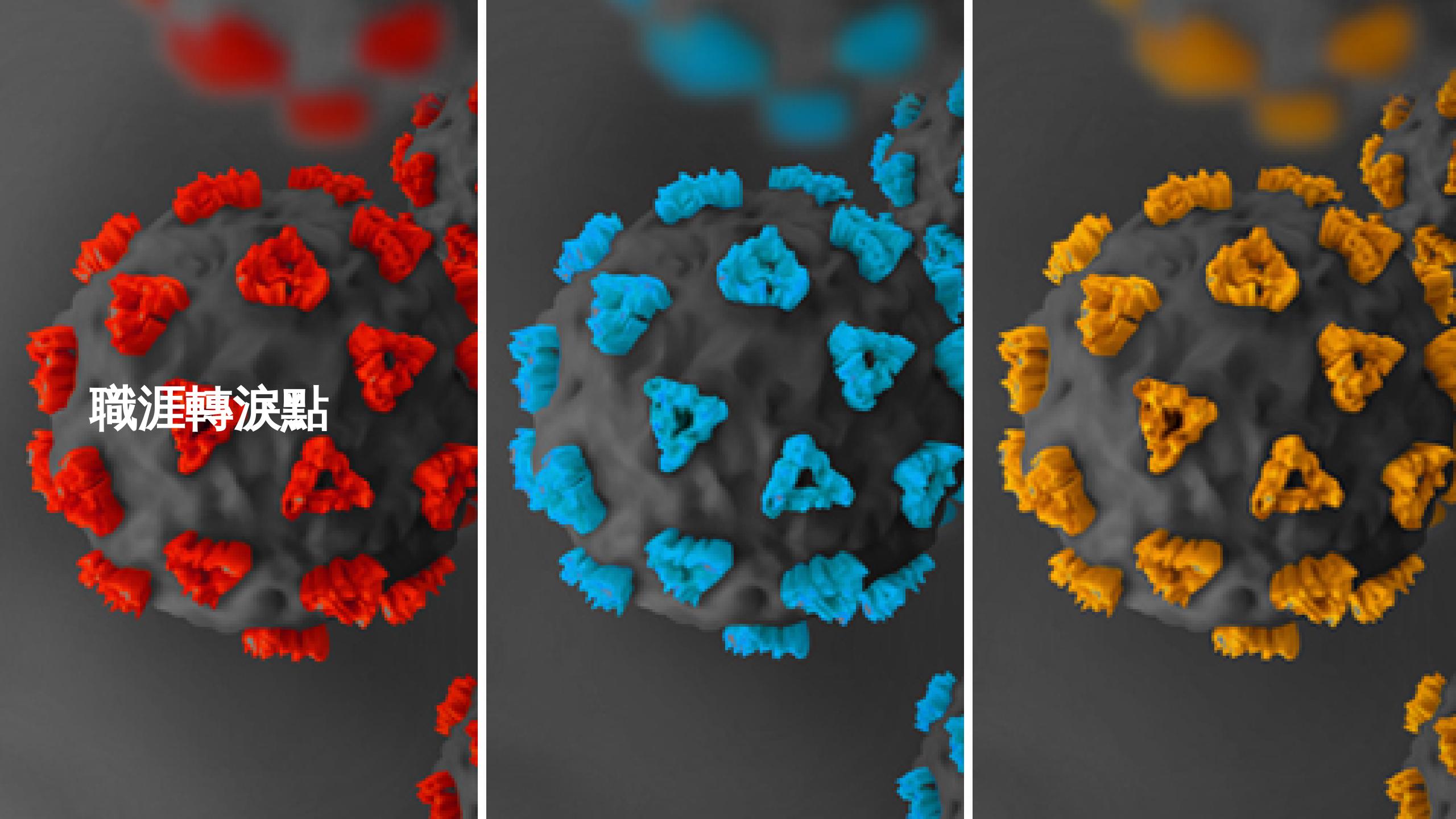
PDF(1712 KB) ↓

285 Read ☆ add

Related articles

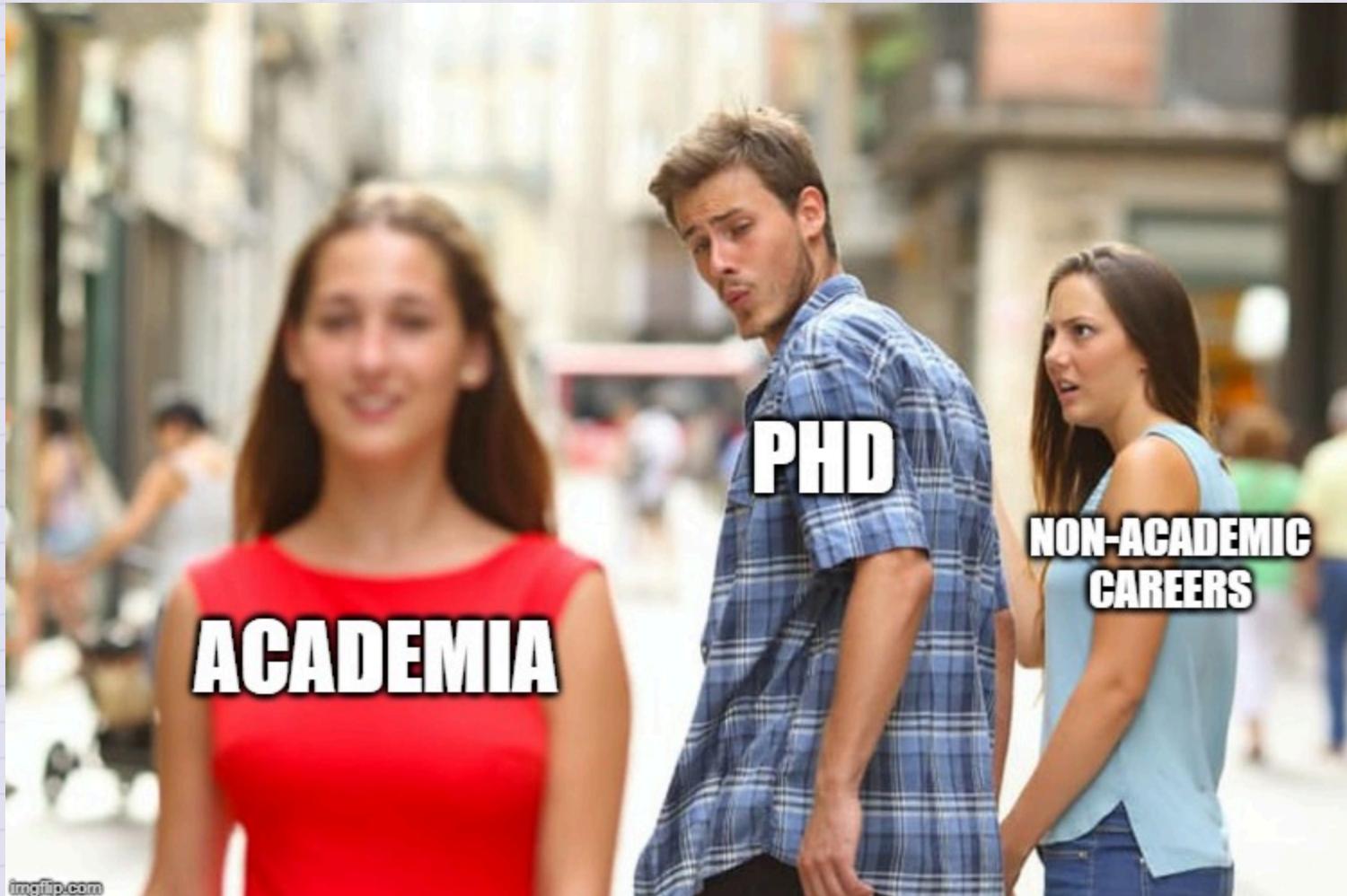
[1] 吴语人称代词复数标记来源的类型学考察

[2] 指人名词人称分裂的语言共性研究

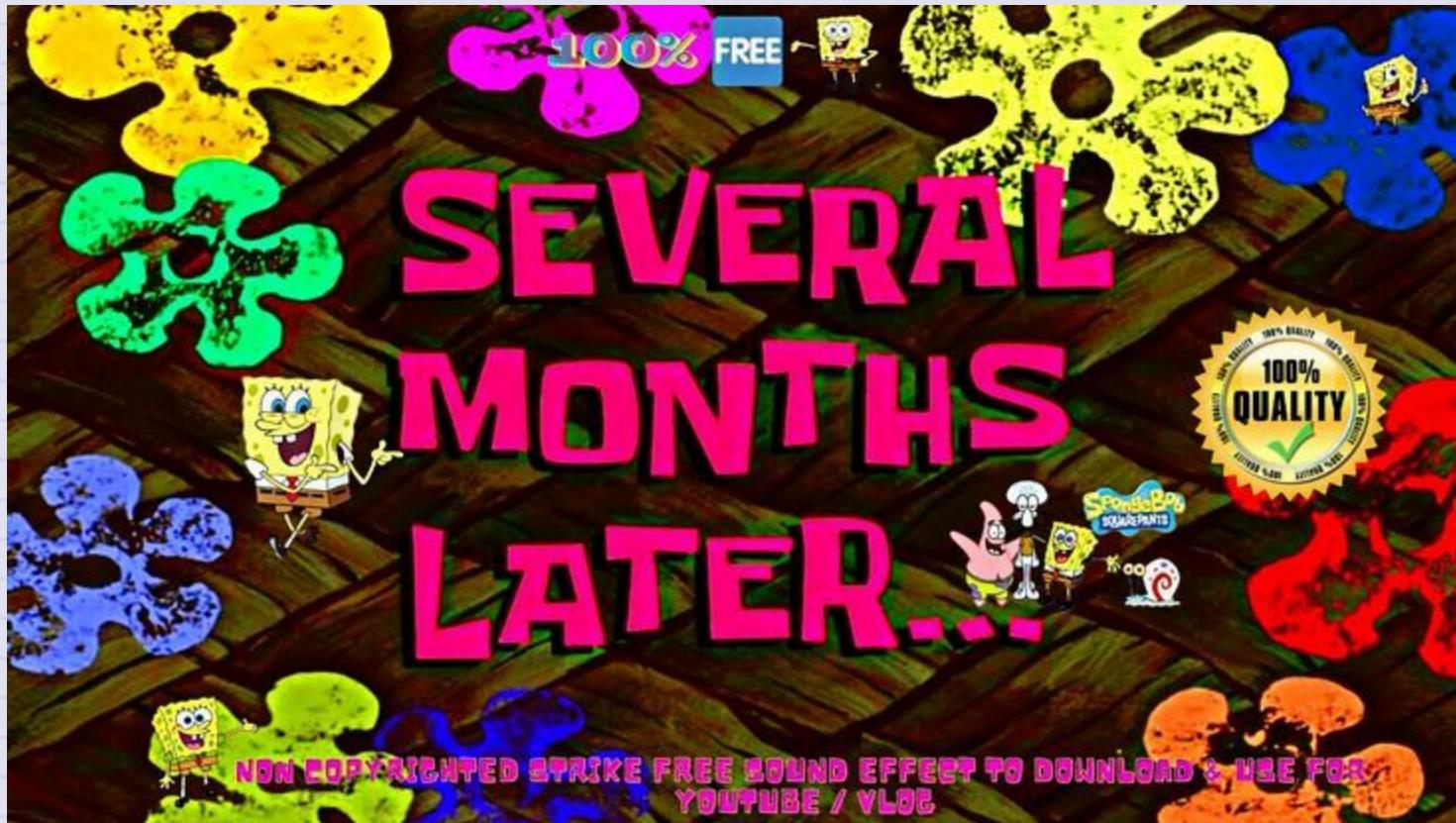


職涯轉捩點

- 心態



- 自學





AI 工程師
哈瑪星科技
創新研發中心

- 業界初體驗



哈瑪星科技
HAMASTAR TECHNOLOGY CO., LTD.



AI 工程師

香港商慧科訊業

AI Lab

- 跨區合作

数据分析
襄理

新光人壽
數位資訊部

- 資安法規



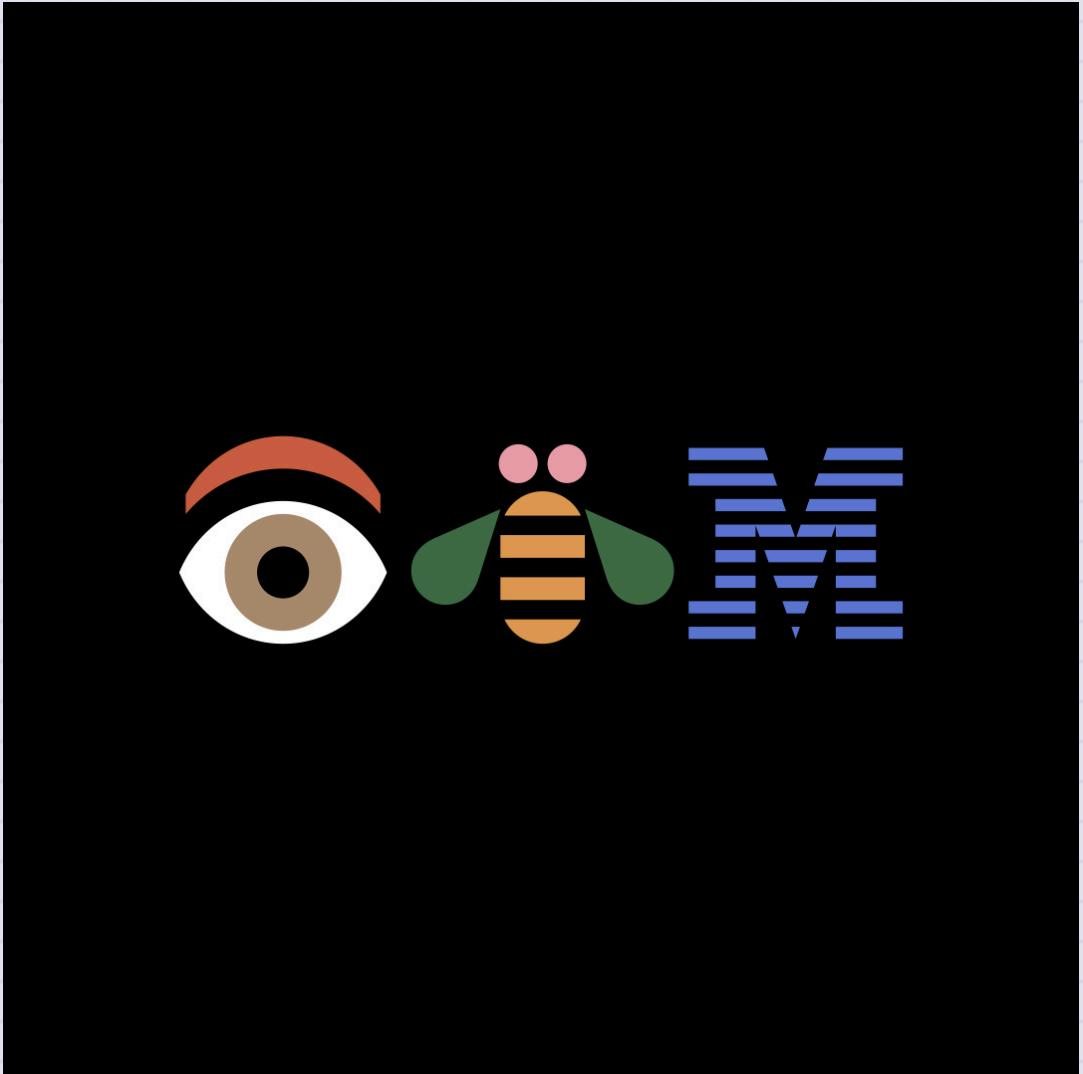
新光金控

新光人壽
Shin Kong Life

AI 暨資料解決
方案技術工程師

IBM
科技事業部

- 美商文化



watsonx: IBM的生成式AI平臺

[source >>](#)



轉職歷程



學習進程

01 Python 程式語言

02 網路爬蟲

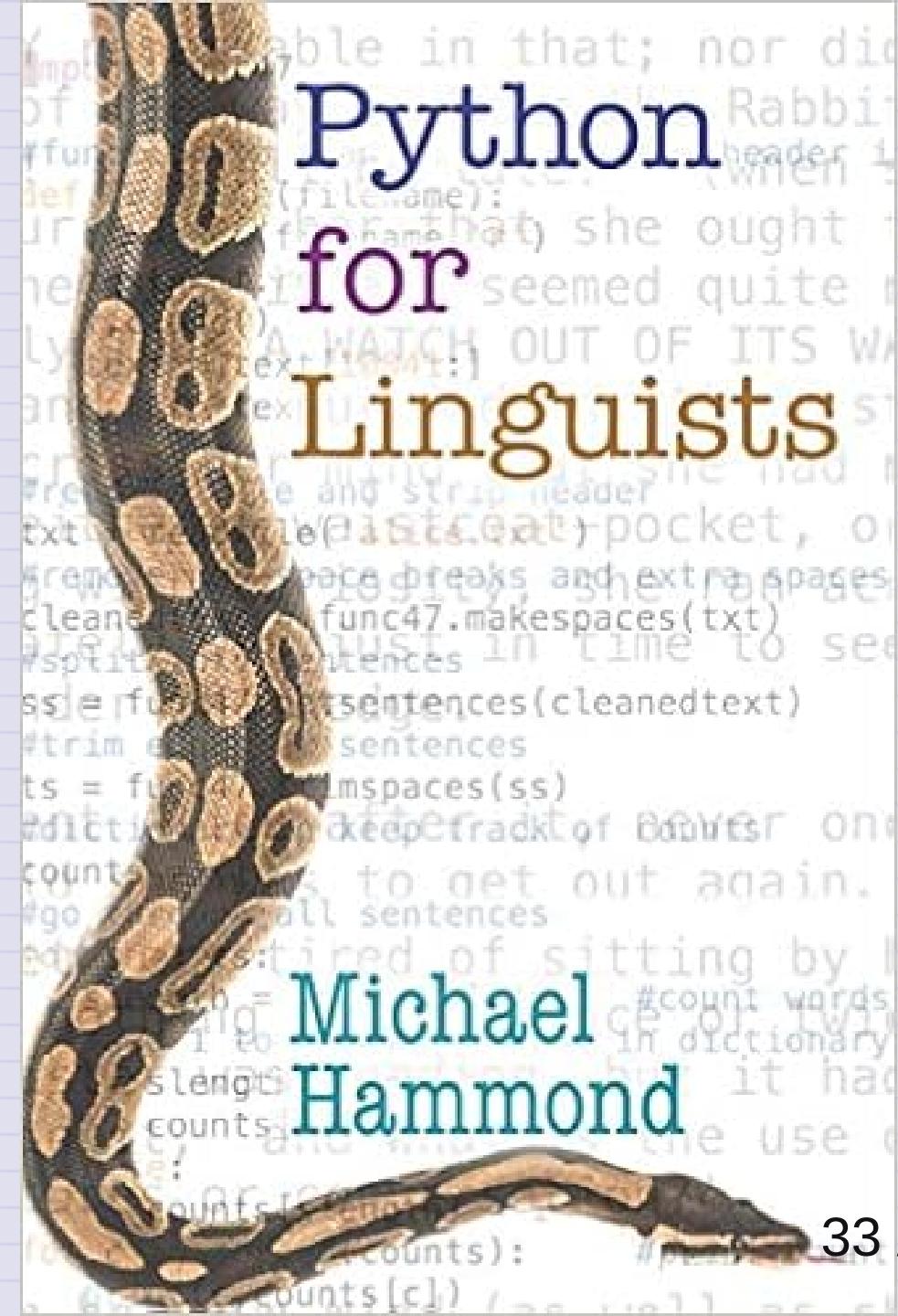
03 自然語言處理

04 機器學習

Python 是蟒蛇也是語言

01 Python 程式語言

語文和工程專業之跨域生涯



程式語言相比自然語言的優勢

- 沒有人是母語者
- 不需練聽說，只需讀寫

"the best way to learn a language
is to speak to natives"
the guy learning Python:



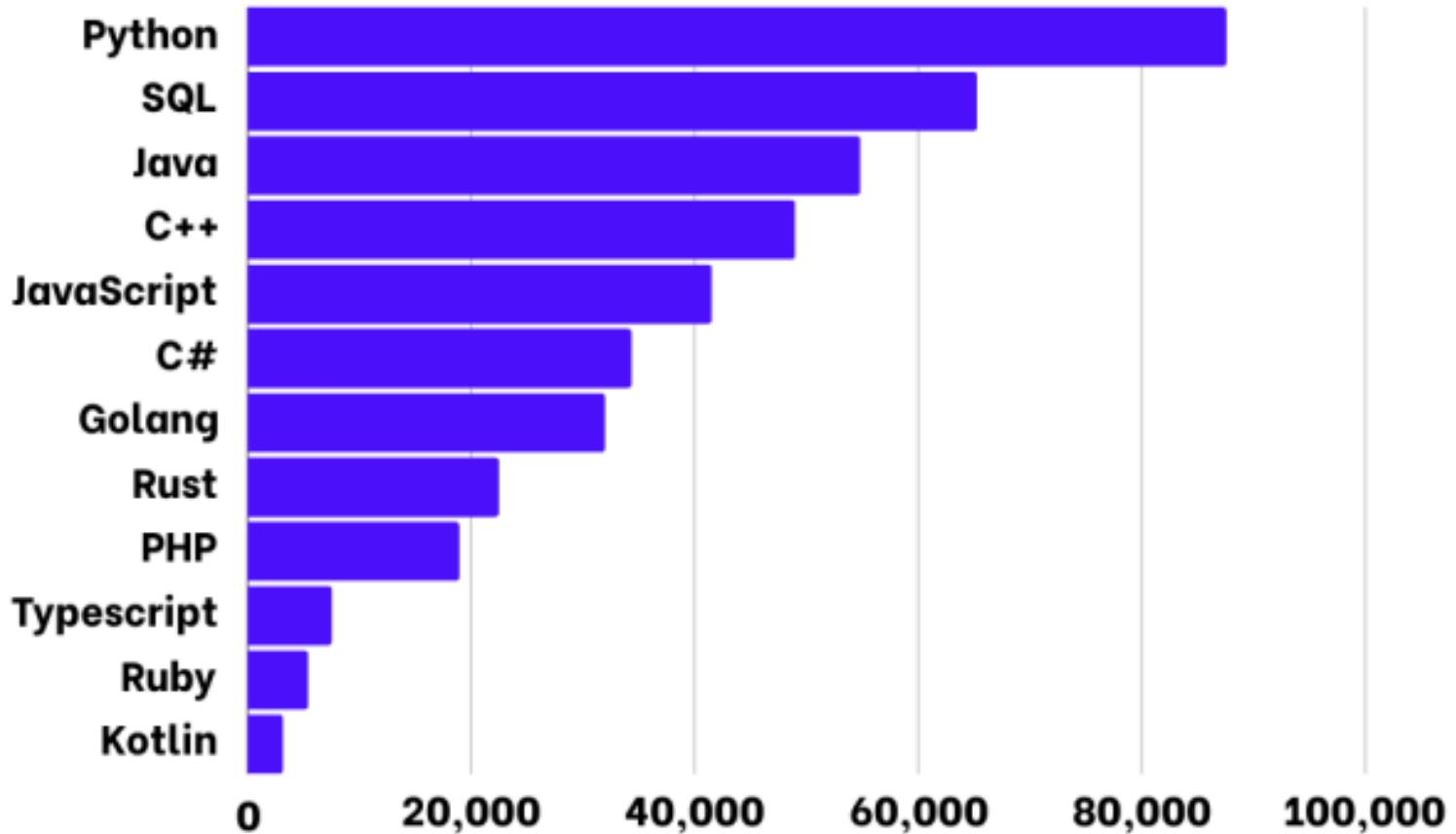
Python 相比其他程式語言的優勢

- 職缺多

[Source >>](#)

2024 Programming Language Breakdown

of open job postings for each language, in the US only



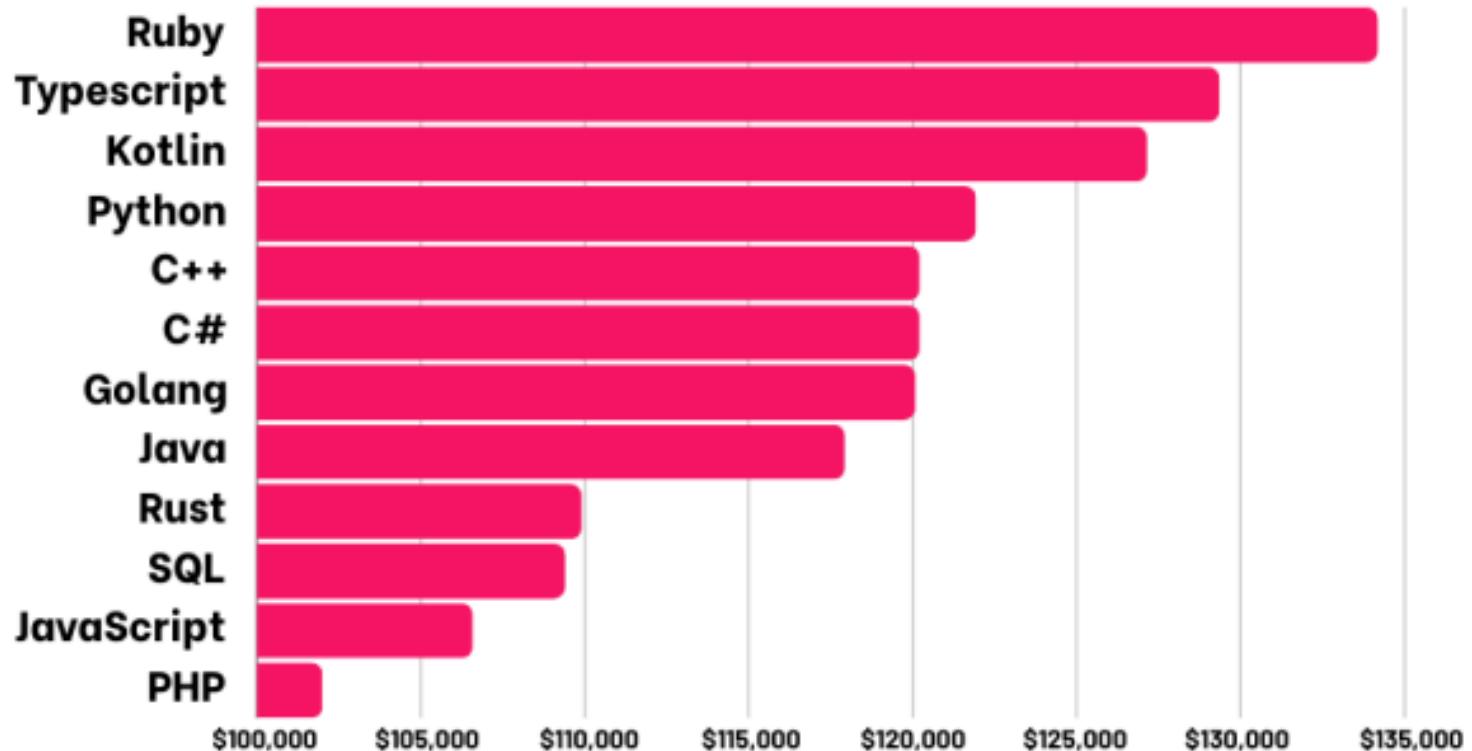
Python 相比其他程式語言的優勢

- 待遇高

[source >>](#)

2024 Programming Language Breakdown

Average Salary in USD



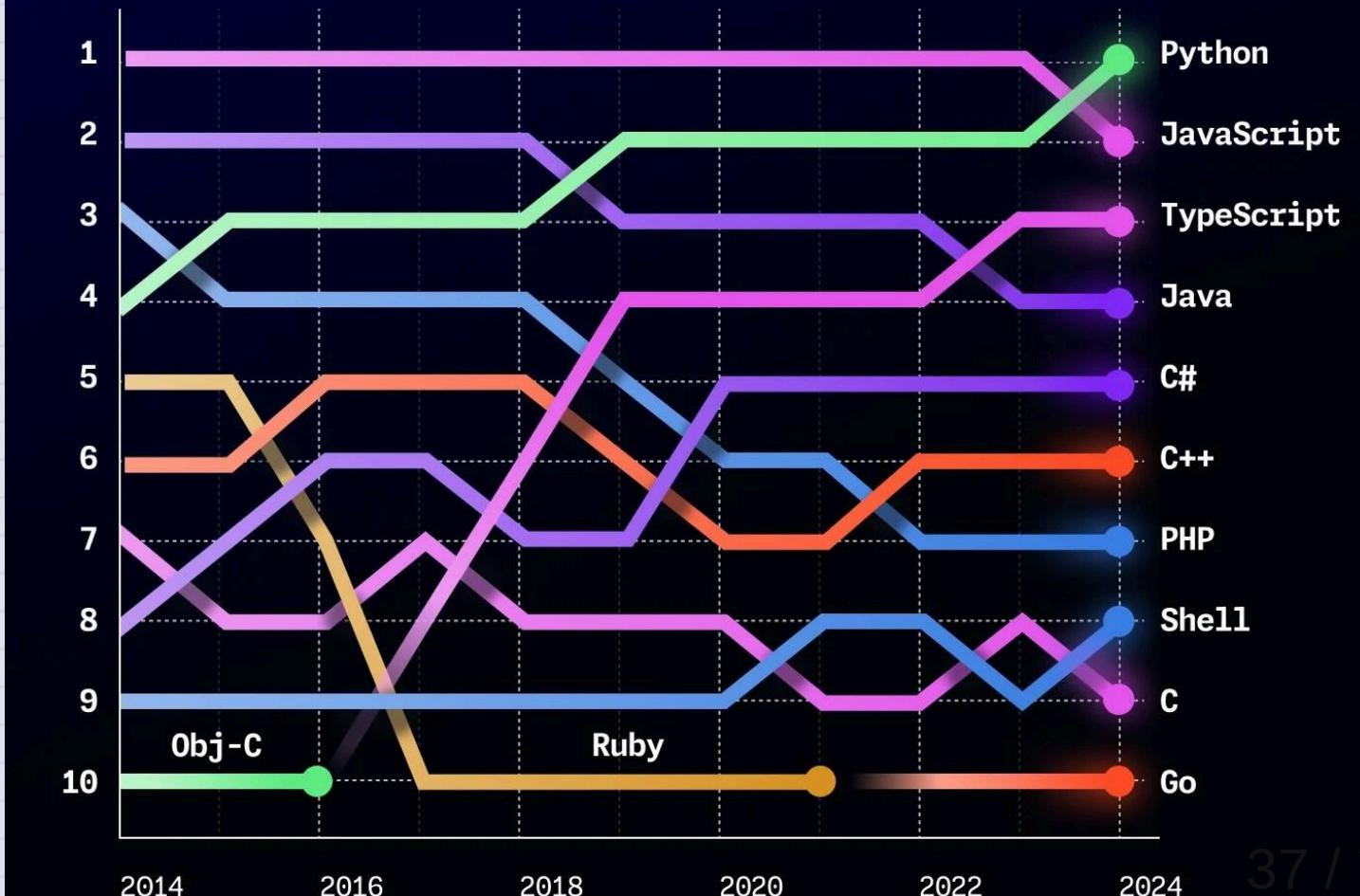
Python在今年
首次成為最熱
門的語言

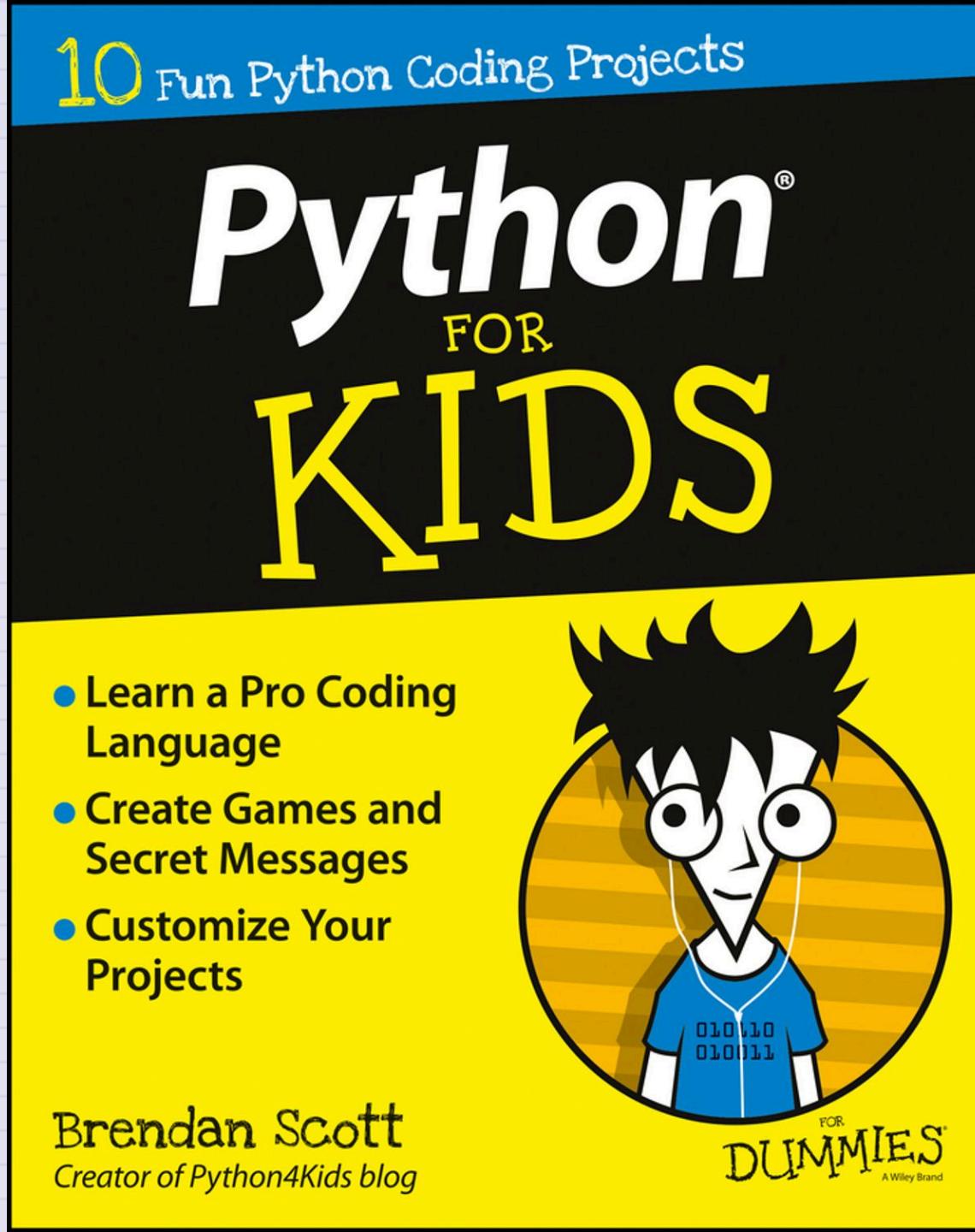
source >>

語文和工程專業之跨域生涯

Top programming languages on GitHub

RANKED BY COUNT OF DISTINCT USERS CONTRIBUTING TO PROJECTS OF EACH LANGUAGE.





連小朋友都學得會!!!
真心不騙！！！

CEFR levels	Study hours	<i>from</i>	<i>to</i>
C2		1,000	1,200
C1		700	800
B2		500	600
B1		350	400
A2		180	200
A1		90	100

The advertisement for the IntelliPaat Python Complete Course features a dark blue background with a grid pattern. At the top left is the IntelliPaat logo. To its right is a white circle containing a blue and yellow Python logo. Below this, the word "PYTHON" is written in large, bold, yellow capital letters. Underneath it, "COMPLETE COURSE" is written in white capital letters. A yellow banner across the middle contains the text "BEGINNER TO ADVANCE" in black and "7 HOURS+" in white. On the right side, a young man with dark hair, wearing a black t-shirt with the IntelliPaat logo, stands looking towards the camera.

軟體工程師 的自學比例

[source >>](#)

90 %

of respondents reported
that they were at least
partially self-taught.

(2017 Stack Overflow developer survey)

高達 80% 的
人透過線上
資源自學

source >>



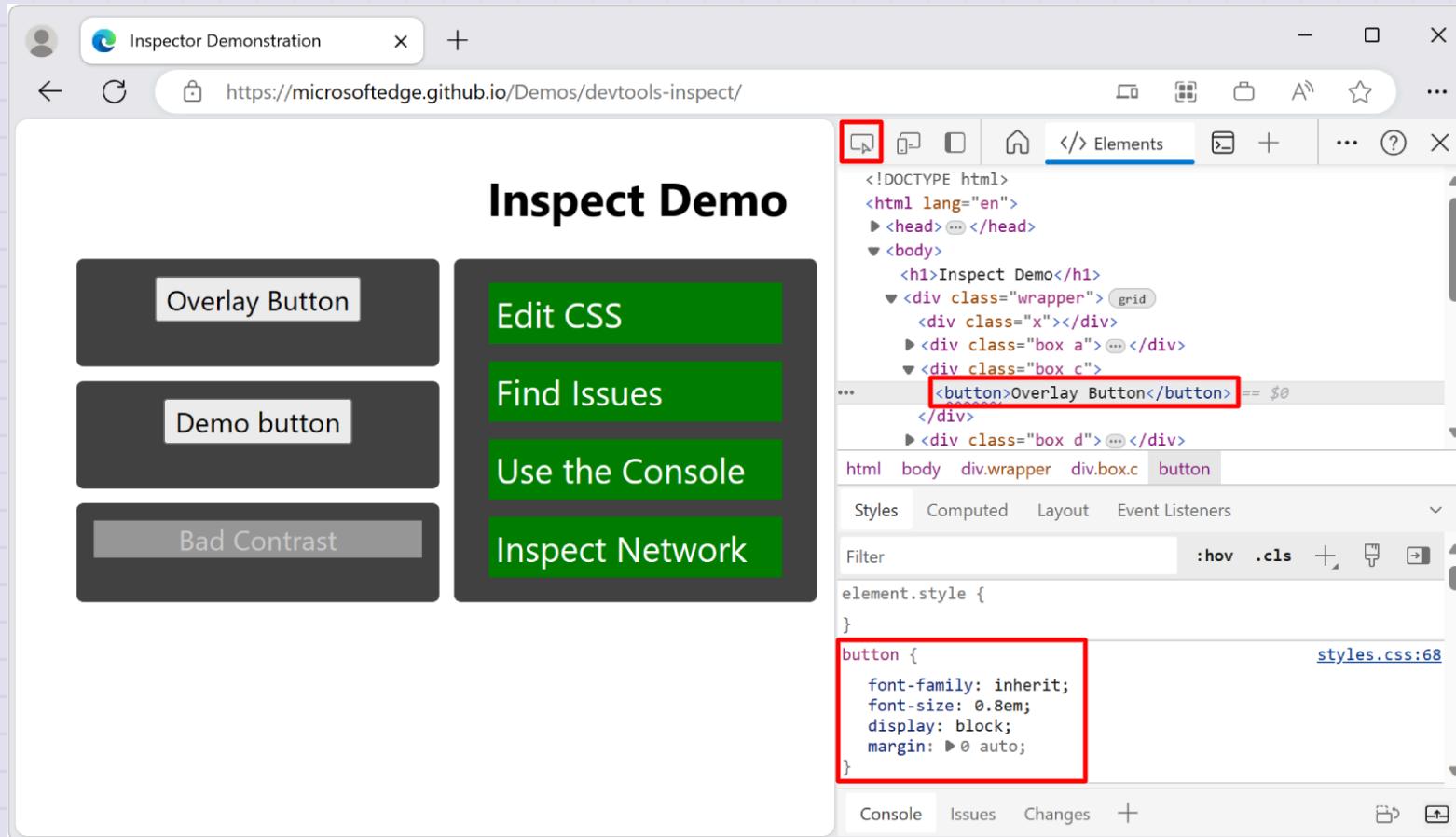
爬蟲能快速取得資料

02 網路爬蟲

Web Scraping Using Python



網路爬蟲 - 用 Python 解析 HTML



契機



族語 E 樂園 >>



代名詞
指示詞



疑問詞



抽象名詞

2024-12-09@NFU

目的地或飯店名稱
高雄
入住日期
2021年05月15日 週六 ~ 2晚
退房日期
2021年05月17日 週一
房間數及人數
1間房，2位成人，0位孩童
搜尋



共找到 252 家高雄飯店

自訂篩選

熱門篩選

- 免費取消
- 即時確認
- 免費WiFi
- 免費早餐
- 很棒：評分4.7以上

星級

- 五星/豪華
- 四星/高級
- 三星/舒適



高雄比歐緻居

Brio Hotel

★★★★★
三多商圈/五福商圈 距離市中心 2公里
限时优惠

很棒 4.7

173則評鑑

-14%

TWD 3,510

TWD 2,984 起

平均每晚含稅價

訂購



和逸飯店(高雄中山館)

Hotel Cozzi Zhongshan Kaohsiung

★★★★★
動人夜景
三多商圈/五福商圈 距離市中心 3公里

很棒 4.7

2,778則評鑑

-14%

TWD 2,874 起

平均每晚含稅價

訂購

1 max_df.head()

ZH_Name	EN_Name	Hotel_ID	City	Location	Address	Campaign	Mask	Hold_Room	Min_Price
0 高雄比歐緻居	Brio Hotel	22260305	高雄	None	中山一路 14-26號	None	None	2	2584
1 和逸飯店 (高雄中山館)	Hotel Cozzi Zhongshan Kaohsiung	2985883	高雄	None	中山二路 260號30樓	None	None	19	2488
2 高雄喜達絲飯店	The Ambassador Hotel Kaohsiung	2328586	高雄	None	成功一路 286號	None	None	3	3306
3 高雄國賓大飯店	The Lees Hotel	701744	高雄	None	民生二路 202號	None	None	9999	2097
4 高雄麗尊酒店	The Lees Hotel	701698	高雄	None	五福一路 105號	None	None	10	2208
5 高雄水京棧國際酒店	H2O Hotel	8194467	高雄	None	明華路 366號	None	None	1	4701

台灣南島語爬蟲成果

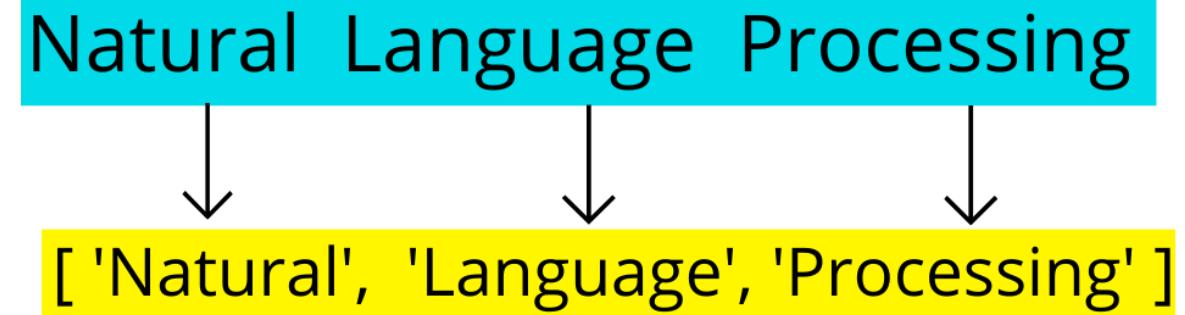
[台灣南島語-華語句庫資料集 APP >>](#)

Lang_En	Lang_Ch	Ab	Ch	From
58650	Saaroa 拉阿魯哇	umacʉka maivʉvʉrʉnga tumuakia umarua sipaasʉrʉpa masʉrʉpʉ.	早上起床，都要用毛巾洗臉。	詞典
115515	Bunun 布農_郡群	Namaun a uvavazan mas hangvang tu balang.	孩子們要吃牛排。	生活會話
92691	Saisiyat 賽夏	rohaenan panrahi: ka rarahi: kayzaeh 'okay tango:i.	晚上掛蚊帳才不會被蚊子叮。	詞典
28392	Rukai 魯凱_霧台	malra patuelre kwini ya cukudhane kwi belebele.	那桌上的香蕉全部拿去。	詞典
117888	Paiwan 排灣_中	kasi nu sun a na kitulu tua pinaiwanan ?	你在哪裡學排灣話的？	九階教材
126445	Puyuma 卑南_南王	munuma tu ami kanu wadi na babayan?	你妹妹幾歲？	句型
6699	Amis 阿美_中部	O cisafakay ko felac no miso.	你的米堆摻有空殼米。	詞典
119373	Rukai 魯凱_東	kadro ka mabocoko si katha, takosiya sa'ararake .	不但吃得飽，並且都很營養。	九階教材
46746	Seediq 賽德克_德固達雅	Mkela melux ka tama mu ma; ini kela smberih.	我父親會對果樹及瓜果施咒；但不會解咒。	詞典
5051	Sakizaya 撒奇萊雅	maudad tu saupuen ku kasuy i salining.	下雨了把木材堆在走廊。	詞典

自然語言處理能透過程式
處理大量文字

03 自然語言處理

Tokenization



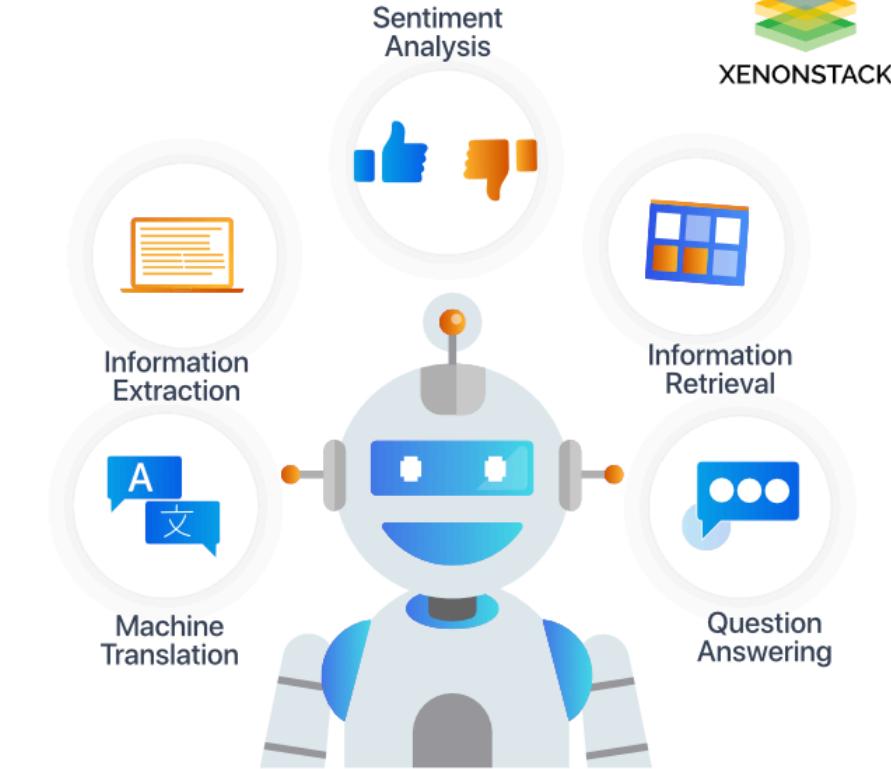
自然語言處理的定義

Natural language processing (NLP) is a subfield of **linguistics**, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to **program computers to process and analyze large amounts of natural language data**.

[Wikipedia >>](#)

NLP 的應用

Applications of **Natural Language Processing** in Different Domains

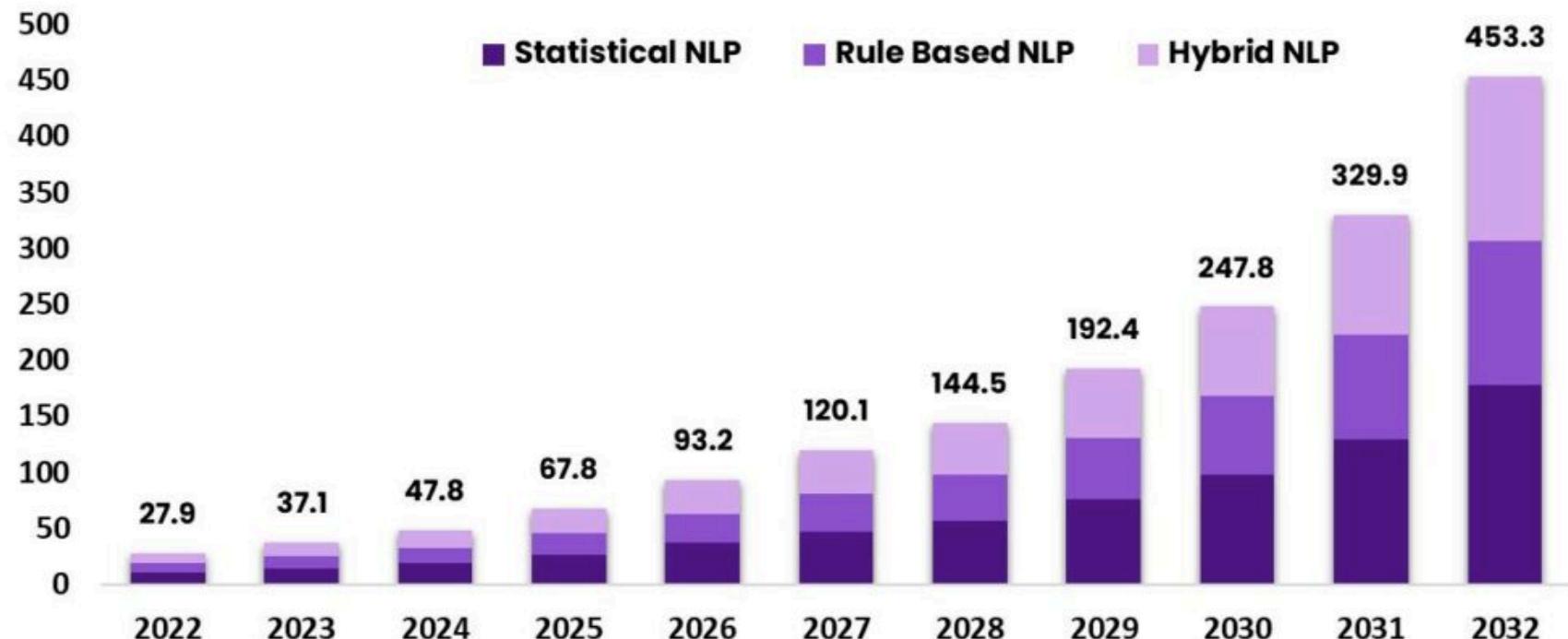


NLP 的市值

source >>

Global Natural Language Processing Market

Size, by type, 2022-2032 (USD Billion)



The Market will Grow
At the CAGR of:

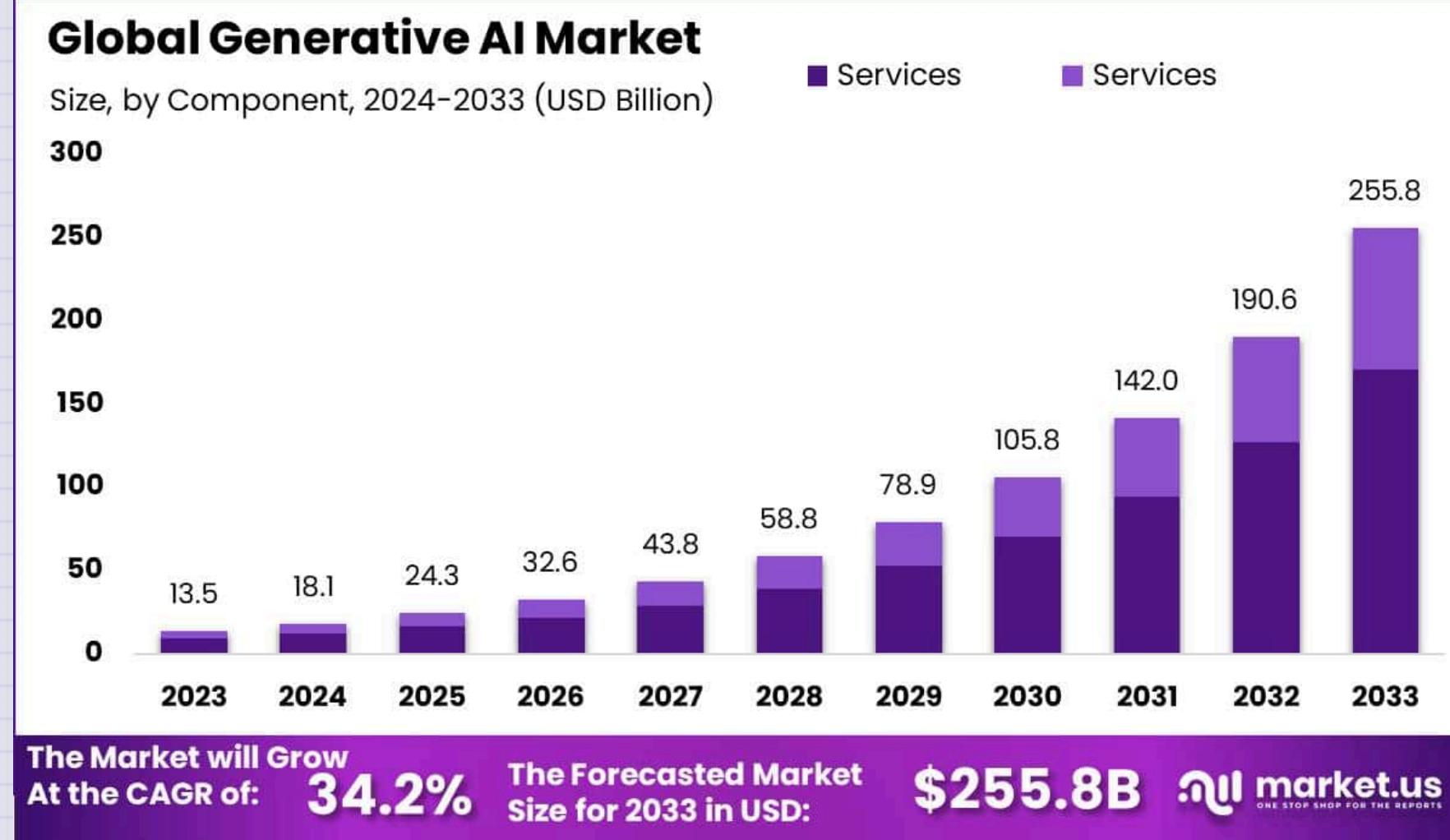
33.1%

The forecasted market
size for 2032 in USD:

\$453.3B  market.us
ONE STOP SHOP FOR THE REPORTS

生成式 AI 的市值

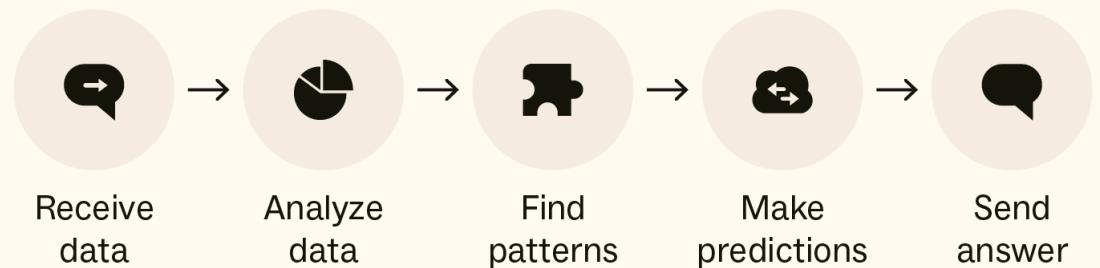
source >>



機器學習能透過大量資料
學習到模式，並做出決策

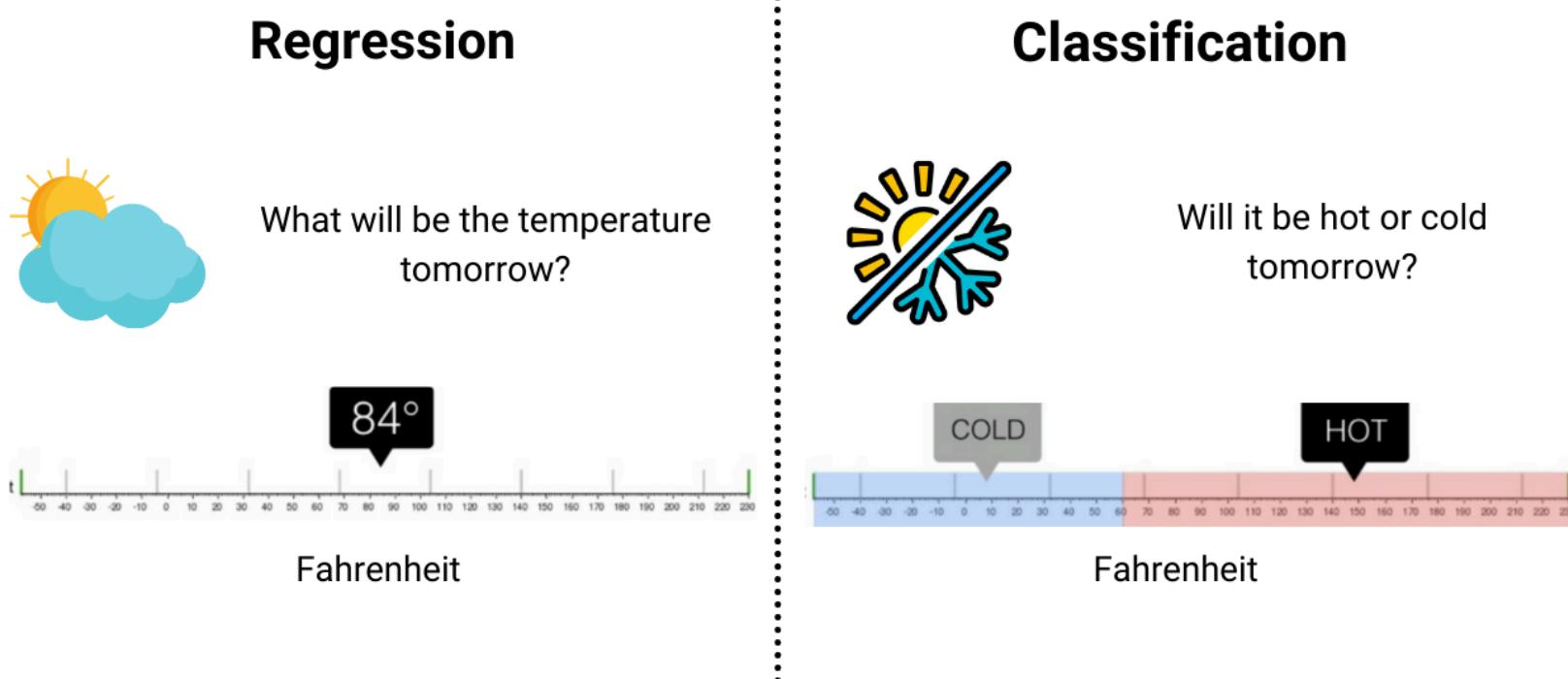
04 機器學習

The machine learning process



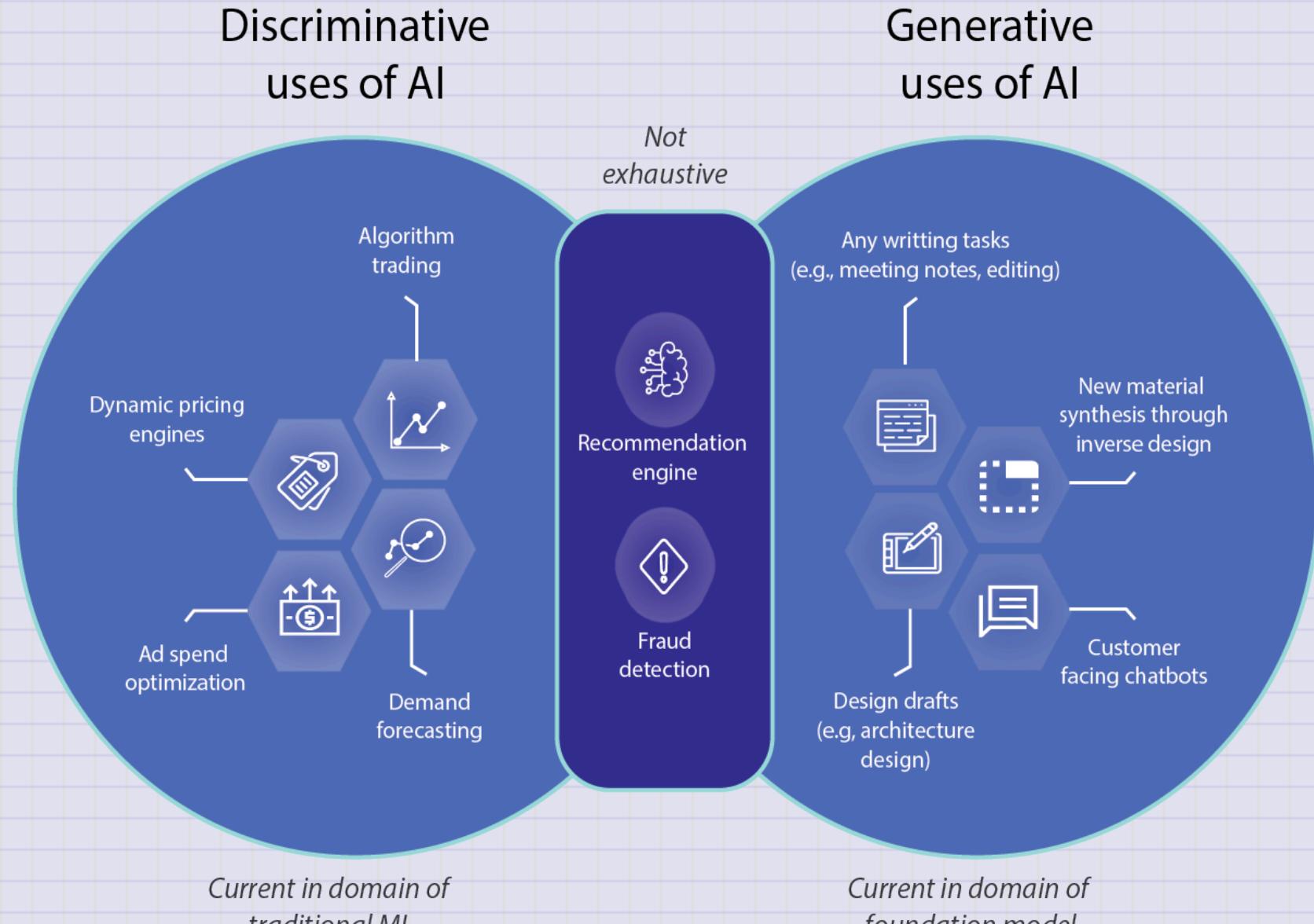
兩類預測模型

- 回歸：
預測數值
- 分類：
預測類別

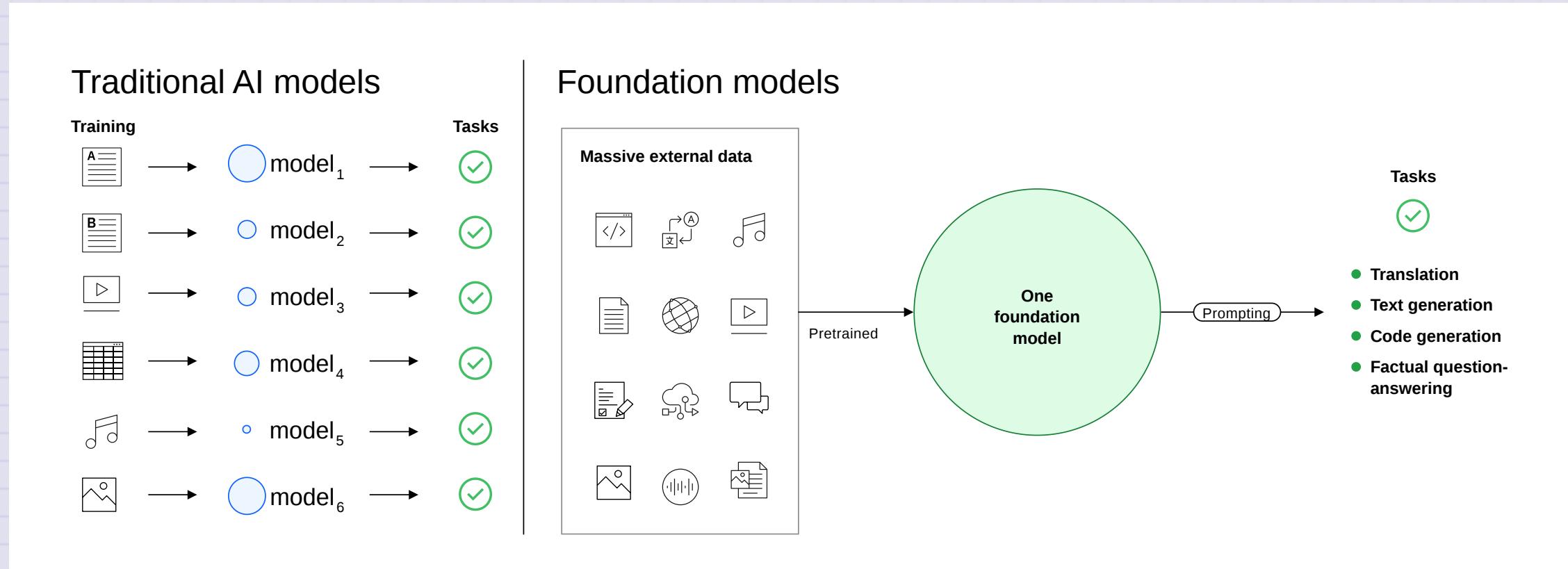


兩類 AI

- 分辨類：
預測數值或
類別
- 生成類：
產生內容

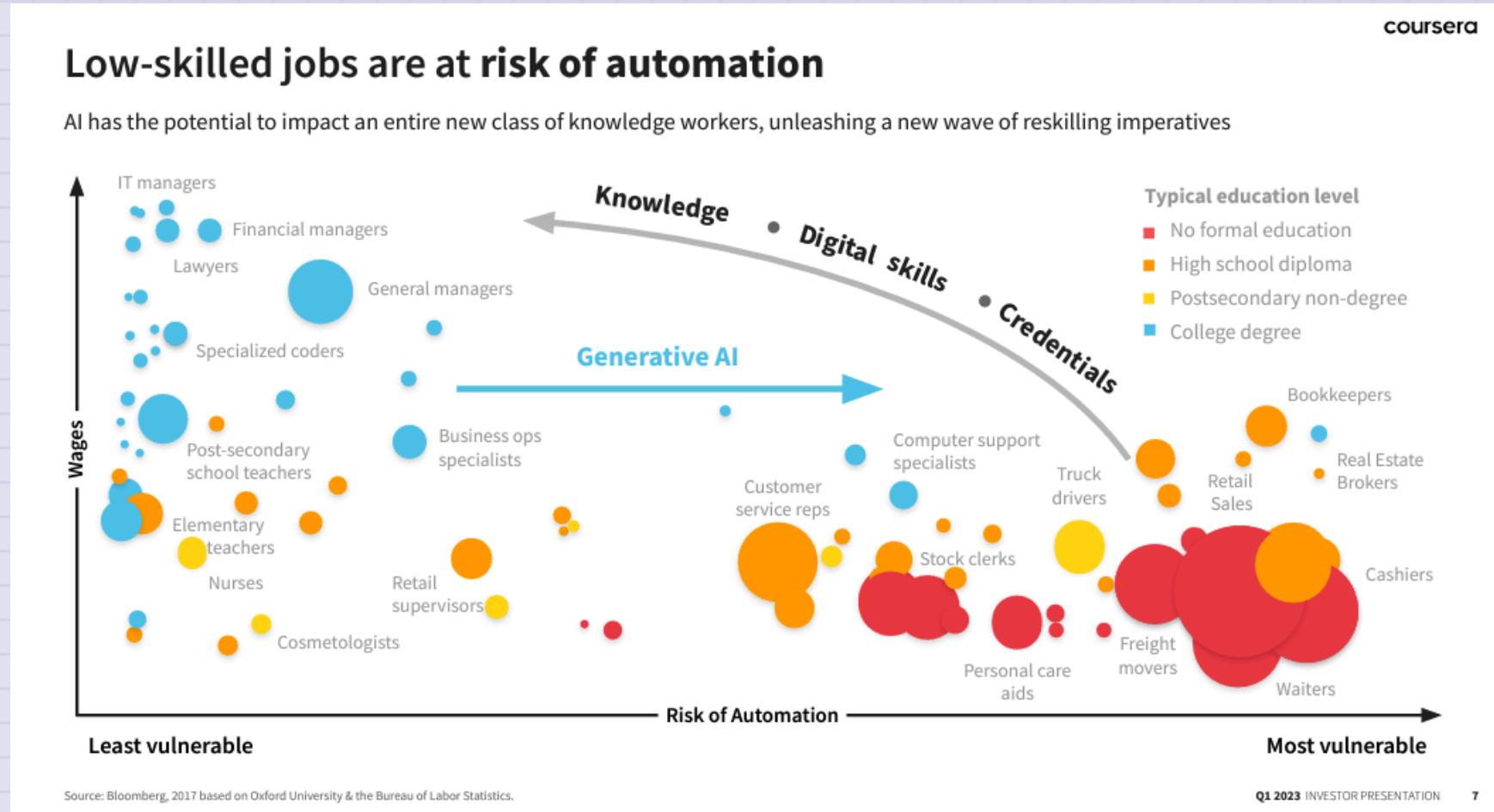


傳統AI與生成式AI



[>> source](#)

生成式 AI 對不 同職業 的衝擊



語文和工程專業之跨域
生涯

學習 4B

01Basics

02Badges

03Bootstrapping

04Blogging

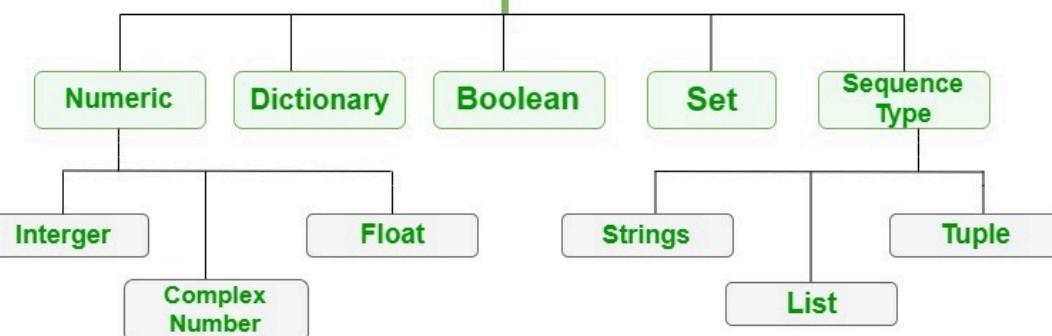
利用geeksforgeeks 一步一步練基本功

01Basics

GeeksforGeeks

A computer science portal for geeks

Python - Data Types



利用cognitiveclass.ai
或IBMSkillsBuild獲取技
術證照

02Badges



COGNITIVE
CLASS.ai



利用GitHub 借力使力開發應用程式

03Bootstrapping

- AI 模型輔助語言學習
[APP >>](#)

語文和工程專業之跨域生涯

Welcome to Streamlit!

Edit [/streamlit_app.py](#) to customize this app to your heart's desire ❤️

If you have any questions, checkout our [documentation](#) and [community forums](#).

In the meantime, below is an example of what you can do with just a few lines of code:

Number of points in spiral

1

2000

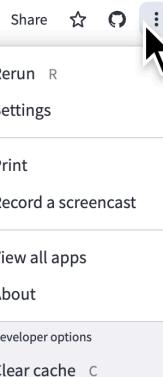
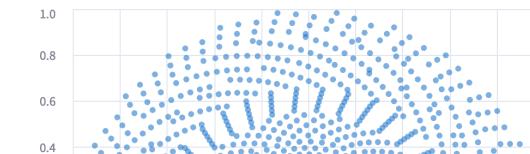
5000

Number of turns in spiral

1

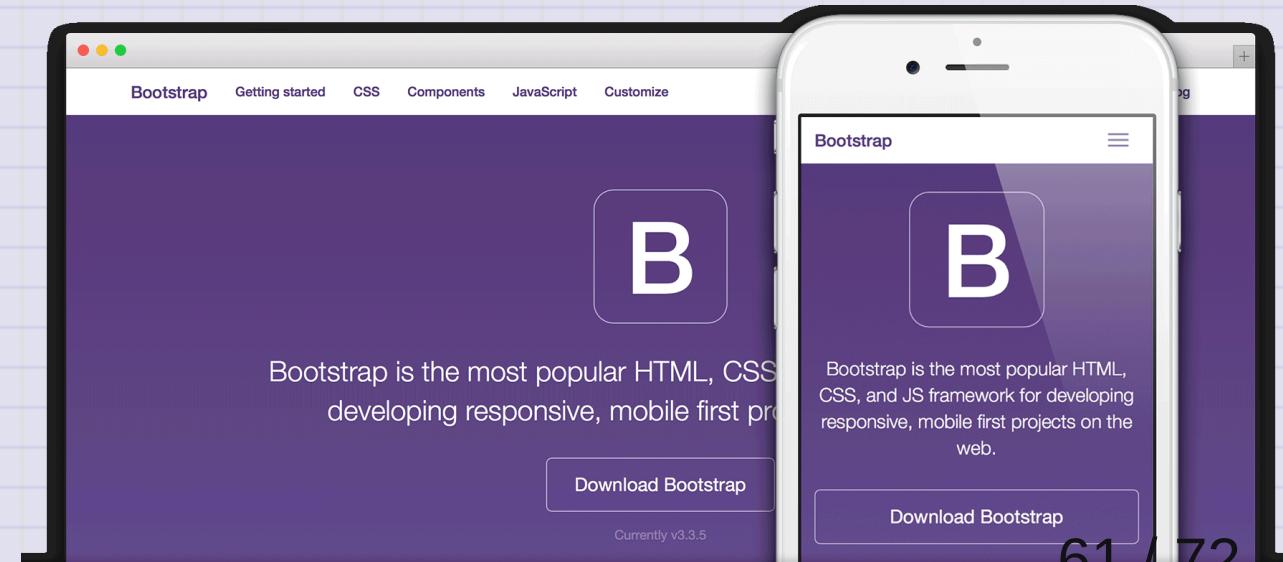
9

100



利用GitHub Pages 撰寫部落格文章分享所學

04Blogging



學位的 投資報酬率 (ROI)

source >>

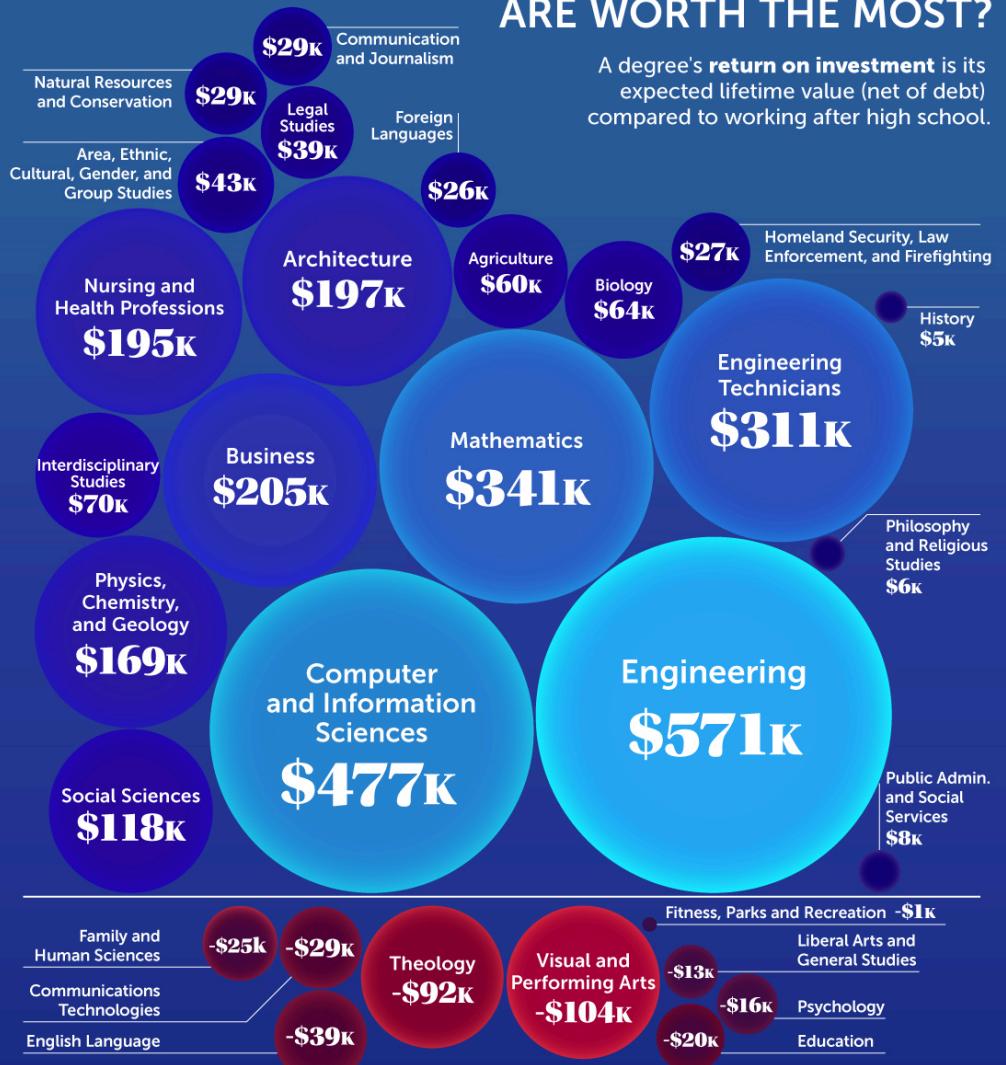
語文和工程專業之跨域生涯

Which Degrees



ARE WORTH THE MOST?

A degree's **return on investment** is its expected lifetime value (net of debt) compared to working after high school.



Source: CollegeNPV

2024 年 10 大 AI 職位

[source >>](#)

語文和工程專業之跨域生涯

10 Highest-Paying AI Jobs in 2024

01

Machine Learning (ML) Engineer
design, build, and deploy machine learning (ML) models

03

Natural Language Processing (NLP) Engineer
Specialize in building ML models to understand and process human language

05

Data Scientist
Extract insights from data to solve problems and inform decisions

07

AI Research Scientist
Explore new methods to make AI systems more intelligent and capable

09

Software Engineer
Turn AI models and algorithms into real-world software applications

02

AI Product Manager
Translate cutting-edge AI technology into real-world solutions

04

Big Data Engineer
Design and implement data pipelines, ensuring data security and integrity

06

Computer Vision Engineer
Build systems that enable computers to "see" and interpret visual data

08

Business Development Manager (BDM)
Finds new clients and grows business for AI companies

10

Prompt Engineer
Craft instructions to unlock the potential of large language models

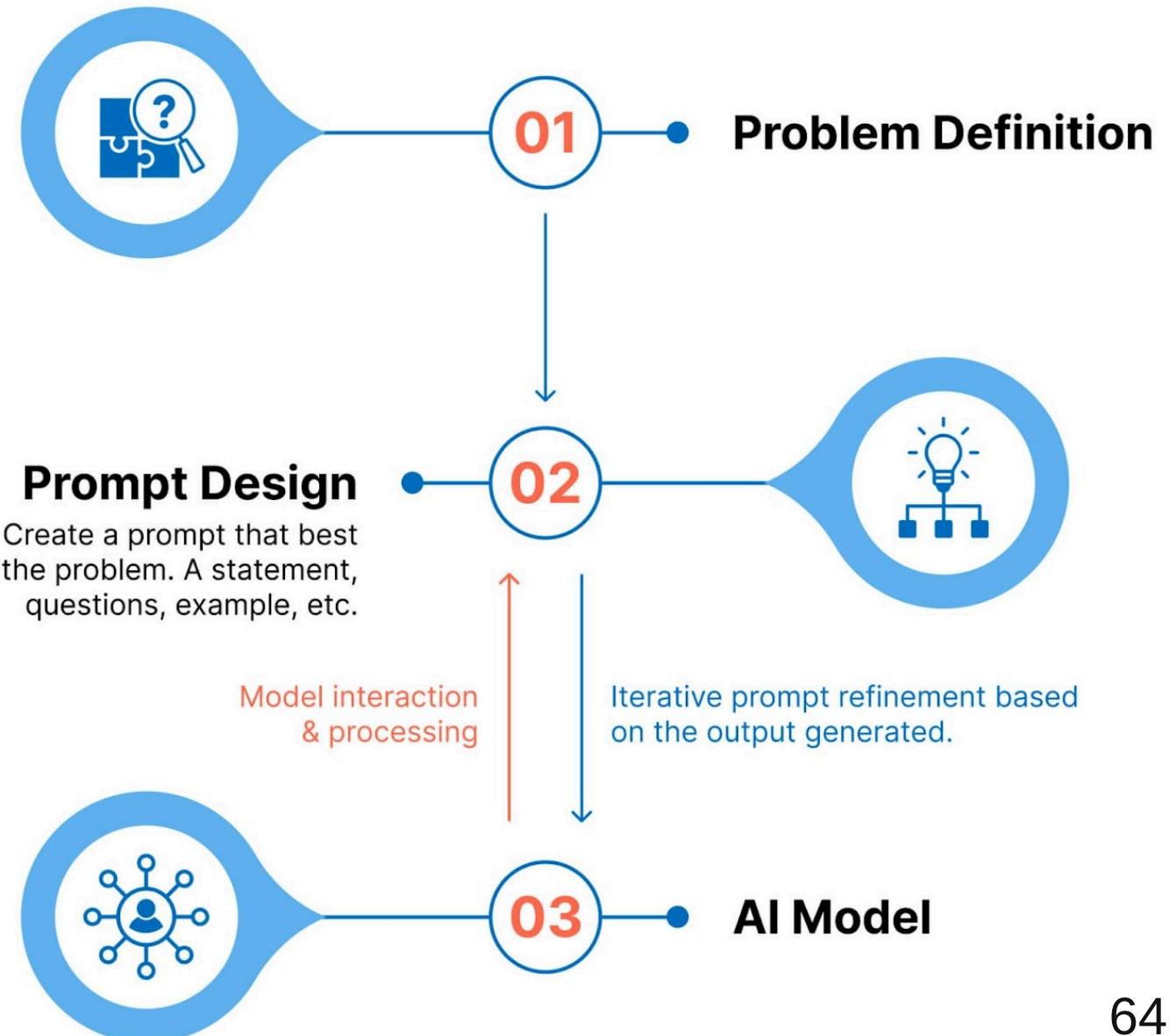
提詞工程

Prompt
Engineering

[source >>](#)

語文和工程專業之跨域生涯

How does Prompt Engineering work?



其實你已經
掌握了最強
的程式語言

[source >>](#)

santiagof.medium.com/english-is-the-most-powerful-programming-language-even-for-data-s... ☆



★ Member-only story

English is the most powerful programming language — even for data science: Introduction to prompt engineering

What prompt engineering is, which are the steps involved in it, and how it changes the way we solve problems with ML.



Facundo Santiago · [Follow](#)

10 min read · May 9, 2023

9 歲小孩可以
你也可以！

source >>



借力LLM

- 常見的大語言模型

source
=>

Notable LLM Makers

Open

Falcon (TII)
Llama 3 (Meta)
Mixtral (Mistral)
Phi 3 (Microsoft)
Qwen 1.5 (Alibaba)
StarCoder 2 (BigCode)



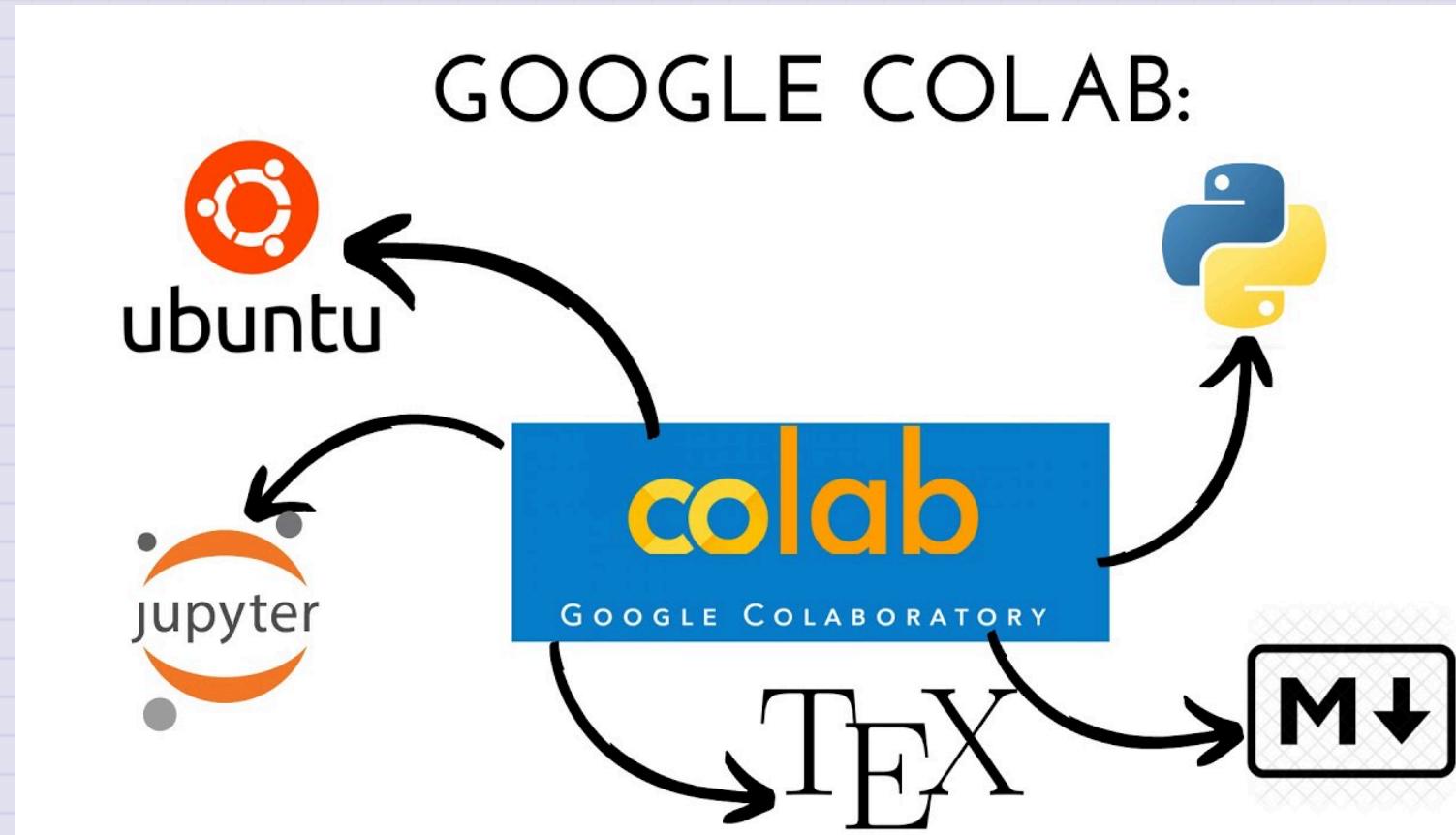
Proprietary

Claude 3 (Anthropic)
Command R+ (Cohere)
GPT 4 (OpenAI)
Gemini 1.5 (Google)
Grok 1.5 (X)
Sonar (Perplexity)



Demo Time!

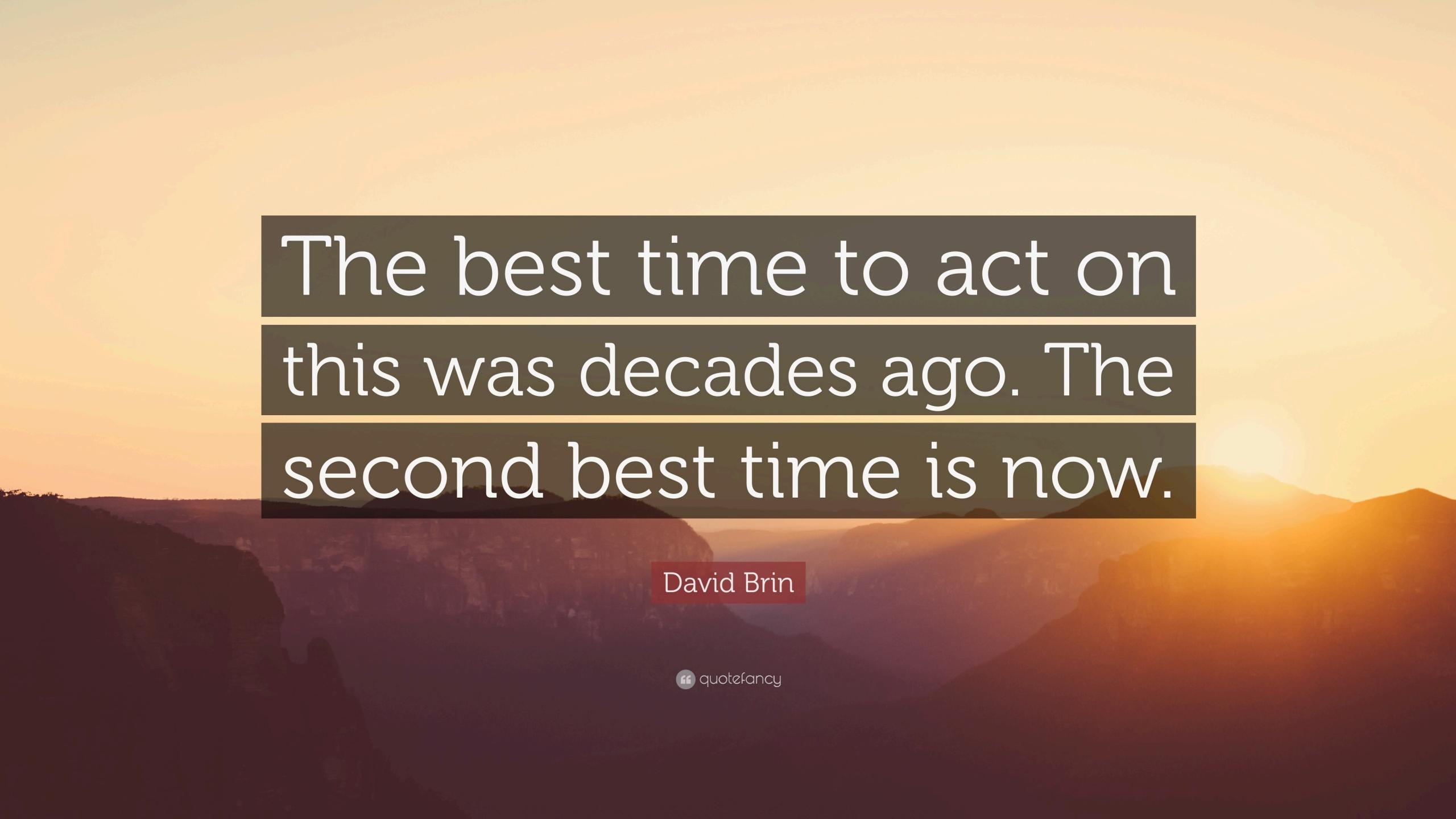
- 點擊這裡進入
Google Colab



結語：有產出的學習更有說服力

- 獲取技術證照
- 撰寫部落格文章
- 開發應用程式





The best time to act on
this was decades ago. The
second best time is now.

David Brin

問答時間？

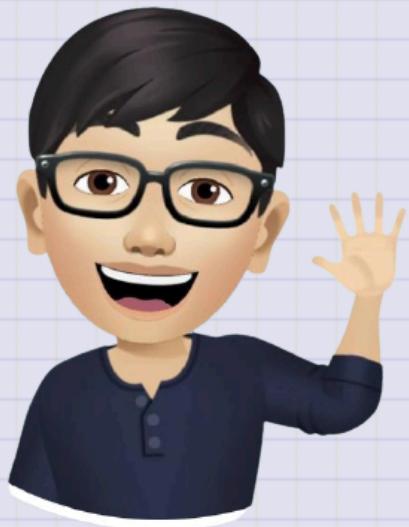


WHAT IF I TOLD YOU

I'M ALLERGIC TO QUESTIONS

MEMEFUL.COM

聯絡方式



- Email:
howard.haowen@gmail.com
- 網站:
<https://howard-haowen.github.io>