

Cloud Study Jam

Generative AI

2023/12/06 (三) 19:00 - 21:00



Speaker: Justin Hsu (許新翎)

- NCCU GDSC - Core Member
- ChainSea - Software Engineer





關於我



Justin Hsu (許新翎)

justin.hsu.1019@gmail.com

- Python Development
- Model Training & Deployment

● 工作 & 社團

- Python 程式家教
- 程曦資訊 軟體工程師 (兼職)
- 政大 GDSC 技術組 核心幹部

● 社群平台

- Website: justin-code.com
- GitHub: [JustinHsu1019](https://github.com/JustinHsu1019)
- LinkedIn: [JustinHsu101999](https://www.linkedin.com/in/JustinHsu101999)
- Instagram: [@Justin.Hsu.99](https://www.instagram.com/Justin.Hsu.99)



關於我



Justin Hsu (許新翎)

justin.hsu.1019@gmail.com

- Python Development
- Model Training & Deployment

專案開發經歷

- 程曦 - AI Trainer 系統
- 程曦 - LLaMA 訓練及部署

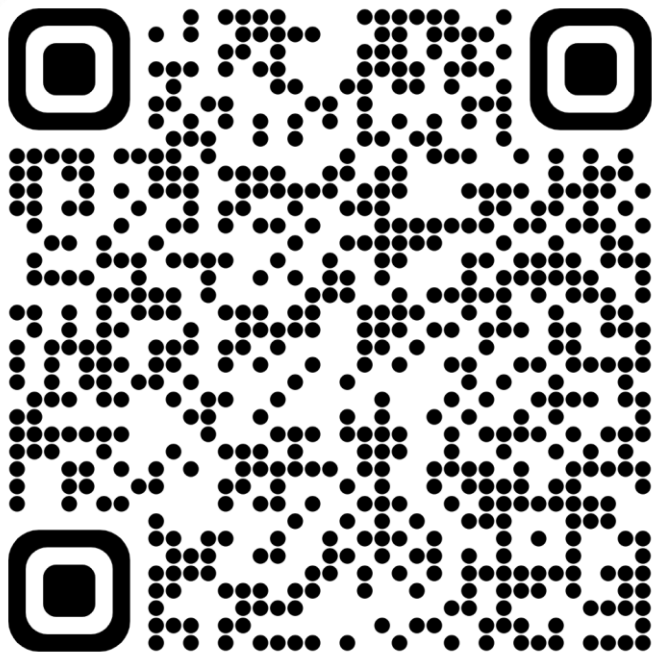
專案開發經歷

- 萊雅 - 客服錄音質檢系統
- 威摩 - 客服知識庫問答系統
- 台電 - 智能機器人優化案



開始之前，拿簡報

<https://bit.ly/GenAIPPT>





活動流程 – 上半場

- 介紹 Generative AI
- 介紹 Google Cloud
- Lab 實作: Detect Labels, Faces, and Landmarks in Images with the Cloud Vision API





活動流程 – 下半場

- 工作坊：Vertex AI PaLM
API：Qwik Start
- 自由討論 + 問問題



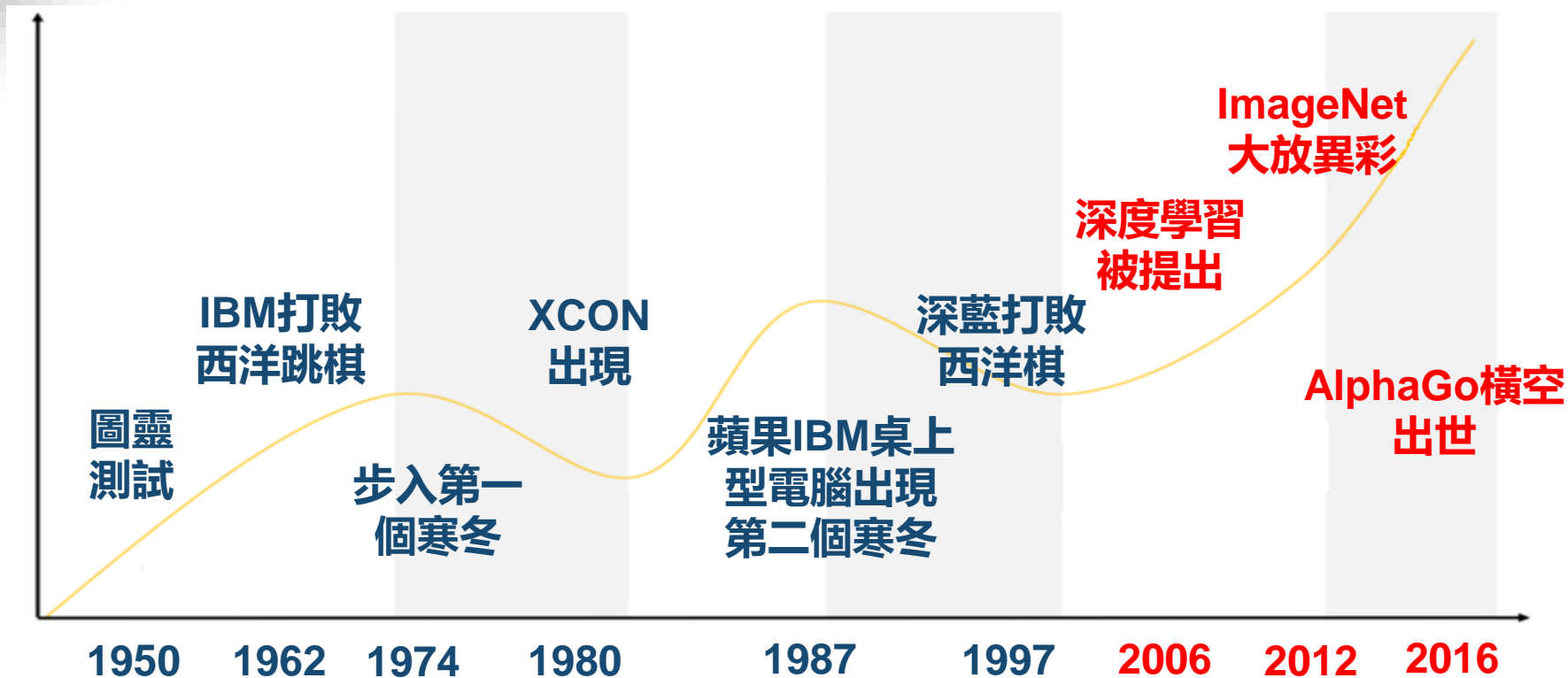


01

生成式AI的發展



AI 的歷史發展





21 世紀 前後AI技術的主要差異

- 21 世紀之前
 - 專家系統為代表
 - 用 “如果-就” (If - Then) 規則來進行推論
 - 規則由「人」來定義，但很多領域「人」無法明確說出規則
- 21 世紀之後
 - 機器學習為代表，其中的深度學習成為主流
 - 主要在解多元方程式： $Y = a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n$
 - 規則由「電腦」自動學習，但需要大量GPU算力



人工智慧的核心

- 機器學習 \approx 解決Y和X之間的函數問題

$$Y = f(x)$$

- Speech Recognition

$$f(\text{audio waveform}) = \text{"How are you"}$$

- Image Recognition

$$f(\text{cat image}) = \text{"Cat"}$$

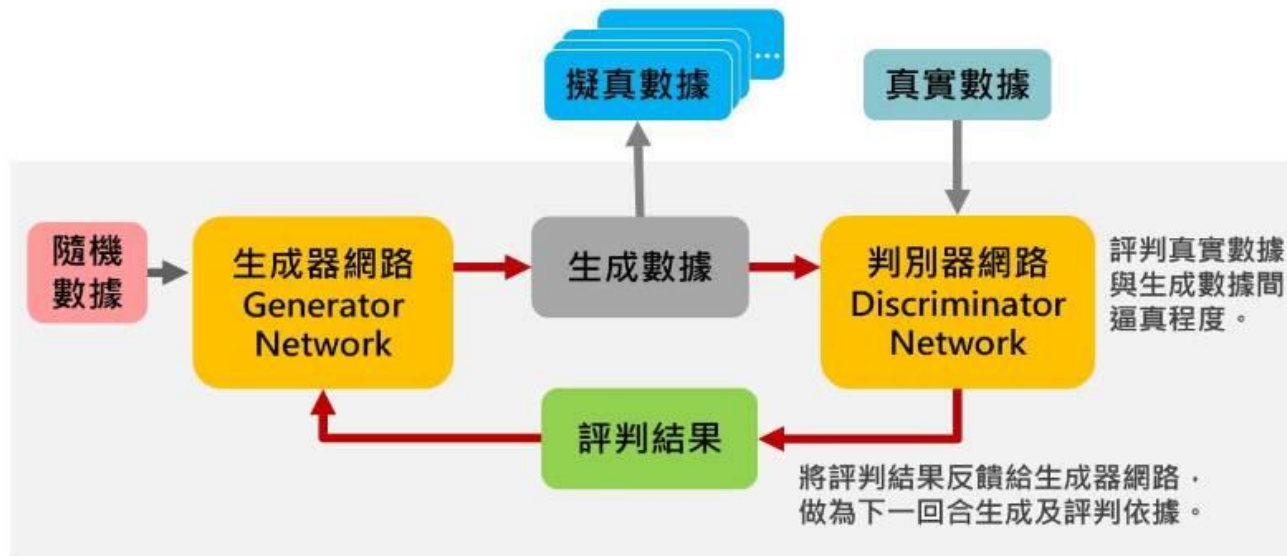


生成式 AI - 生成對抗網路 (GANs)

- 2014年
伊恩·古德費洛等人
提出生成對抗網路

- 生成圖片
- 生成影片
- 生成音樂

GAN架構基本工作原理





生成式 AI - Transformer

- 2017年，Google提出Transformer模型，解決了長序列處理和並行處理的問題，讓處理資料長度和速度得到大幅提昇，成為生成式AI的主流
- 2018年，Google 推出Bert 模型，在多項自然語言處理任務上取得了新的突破
- 同年，OpenAI 推出GPT（G 就是生成，T 就是Transformer）
- 2022年，OpenAI 推出 ChatGPT
- 2023年，Meta 推出 LLaMA
- 2023年，Google 推出 PaLM2 和 Bard



GPT 系列 主要原理

- 類似文字接龍

$$Y = w_1x_1 + w_2x_2 + \dots + w_nx_n$$

- 例：

x_1 是"我"， x_2 是"是"

然後要猜接下來的字 Y 是什麼



02

提示詞設計



"人類不會被AI取代，但善用AI的人，將取代不會用的人。"



如何跟大模型溝通

大模型的輸入：提示詞 Prompt

- 提示詞的基本結構
- 大模型提示詞的5大技巧
- 參數設定



提示詞的基本結構

- 任務：
要大模型做什麼，例如「請總結下面這篇文章」
- 輸入
要大模型處理的資料，例如要總結的文章內容
- 輸出
要大模型輸出的內容，例如總結成怎樣的格式



用好大模型提示詞的5大技巧

- 簡潔不囉唆
- 具體且明確定義
- 一次只做一個任務
- 把開放問題變成選擇題
- 使用Few Shots技巧給予範例

Start Lab

01:00:00

Vertex AI PaLM API: Qwik Start

1 hour

Free



GSP1155



Google Cloud Self-Paced Labs

Text Prompts

The following table shows the parameters that you need to configure for the Vertex AI PaLM API for text:

Parameter	Description	Acceptable values
prompt	Text input to generate model response. Prompts can include preamble, questions, suggestions, instructions, or examples.	Text
temperature	The temperature is used for sampling during the response generation, which occurs when topP and topK are applied. Temperature controls the degree of randomness in token selection. Lower temperatures are good for prompts that require a more deterministic and less open-ended or creative response, while higher temperatures can lead to more diverse or creative results. A temperature of 0 is deterministic: the highest probability response is always selected. For most use cases, try starting with a temperature of 0.2.	0.0–1.0 Default: 0

maxOutputTokens	<p>Maximum number of tokens that can be generated in the response. Specify a lower value for shorter responses and a higher value for longer responses.</p> <p>A token may be smaller than a word. A token is approximately four characters. 100 tokens correspond to roughly 60-80 words.</p>	1–1024 Default: 0
topK	<p>Top-k changes how the model selects tokens for output. A top-k of 1 means the selected token is the most probable among all tokens in the model's vocabulary (also called greedy decoding), while a top-k of 3 means that the next token is selected from among the 3 most probable tokens (using temperature).</p> <p>For each token selection step, the top K tokens with the highest probabilities are sampled. Then tokens are further filtered based on topP with the final token selected using temperature sampling.</p> <p>Specify a lower value for less random responses and a higher value for more random responses.</p>	1–40 Default: 40
topP	<p>Top-p changes how the model selects tokens for output. Tokens are selected from most K (see topK parameter) probable to least until the sum of their probabilities equals the top-p value. For example, if tokens A, B, and C have a probability of 0.3, 0.2, and 0.1 and the top-p value is 0.5, then the model will select either A or B as the next token (using temperature) and doesn't consider C. The default top-p value is 0.95.</p> <p>Specify a lower value for less random responses and a higher value for more random responses.</p>	0.0–1.0 Default: 0.95



Prompt

Parameter	Description	Acceptable values
<code>prompt</code>	Text input to generate model response. Prompts can include preamble, questions, suggestions, instructions, or examples.	Text



Max Output Tokens

`maxOutputTokens`

Maximum number of tokens that can be generated in the response. Specify a lower value for shorter responses and a higher value for longer responses.

1–1024

A token may be smaller than a word. A token is approximately four characters. 100 tokens correspond to roughly 60–80 words.

Default: 0



Top K

<p>topK</p>	<p>Top-k changes how the model selects tokens for output. A top-k of 1 means the selected token is the most probable among all tokens in the model's vocabulary (also called greedy decoding), while a top-k of 3 means that the next token is selected from among the 3 most probable tokens (using temperature).</p> <p>For each token selection step, the top K tokens with the highest probabilities are sampled. Then tokens are further filtered based on topP with the final token selected using temperature sampling.</p> <p>Specify a lower value for less random responses and a higher value for more random responses.</p>	<p>1-40</p> <p>Default: 40</p>
-------------	---	--------------------------------



範例

Text input

The name of that
country is the

Language
Model

Text Output

United

知乎 @abnercloud

Text input

The name of that
country is the



Language Model

United

Text Output

知乎 @abnercloud



挑選 Top K

1) Consider only the top 3 tokens.
Ignore all others.



2) Sample from them based on
their likelihood scores.





Top P

topP

Top-p changes how the model selects tokens for output. Tokens are selected from most K (see topK parameter) probable to least until the sum of their probabilities equals the top-p value. For example, if tokens A, B, and C have a probability of 0.3, 0.2, and 0.1 and the top-p value is 0.5, then the model will select either A or B as the next token (using temperature) and doesn't consider C. The default top-p value is 0.95.

Specify a lower value for less random responses and a higher value for more random responses.

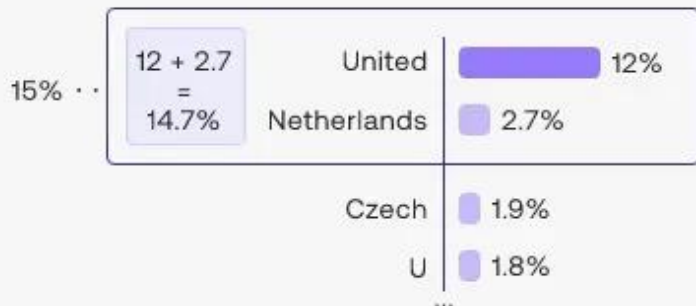
0.0–1.0

Default:
0.95

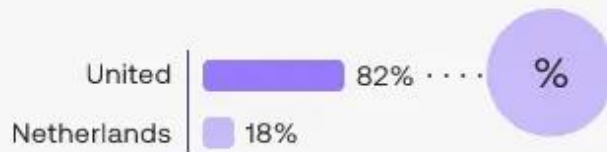


挑選 Top P

1- Consider only the top tokens whose likelihoods add up to 15%. Ignore all others.



2- Sample from them based on their likelihood scores.





Temperature

temperature

The temperature is used for sampling during the response generation, which occurs when `topP` and `topK` are applied. Temperature controls the degree of randomness in token selection. Lower temperatures are good for prompts that require a more deterministic and less open-ended or creative response, while higher temperatures can lead to more diverse or creative results. A temperature of 0 is deterministic: the highest probability response is always selected. For most use cases, try starting with a temperature of 0.2.

0.0–1.0

Default: 0

介紹 Google Cloud





為什麼要學 Google Cloud

Google Cloud 是 Google 將自己的雲端平台基礎建設（如：虛擬機器、網路、儲存空間、資料分析或機器學習服務等等）開放給程式開發者、IT 人員佈建自己的系統或程式

學會 Google Cloud，您可以做到以下幾件事（包含但不限於），例如：

1. **部署自己的網站**，供全世界的用戶存取。
2. 隨時在 Google Cloud 上**建立或刪除虛擬機器**，提供各種測試或開發用途。
3. 使用 Google Cloud 上的 **AI 或 ML 服務**，開發自己的智慧應用程式。
4. 使用 Google Cloud 上的**資料分析工具**，完成大數據的資料處理及分析。



介紹 Cloud Skills Boost





什麼是 Cloud Skills Boost

Cloud Skills Boost 是 Google Cloud 提供的官方培訓和認證平台。這個平台為希望學習和掌握 Google Cloud 產品和服務的個人和團隊提供了一系列的課程和資源。

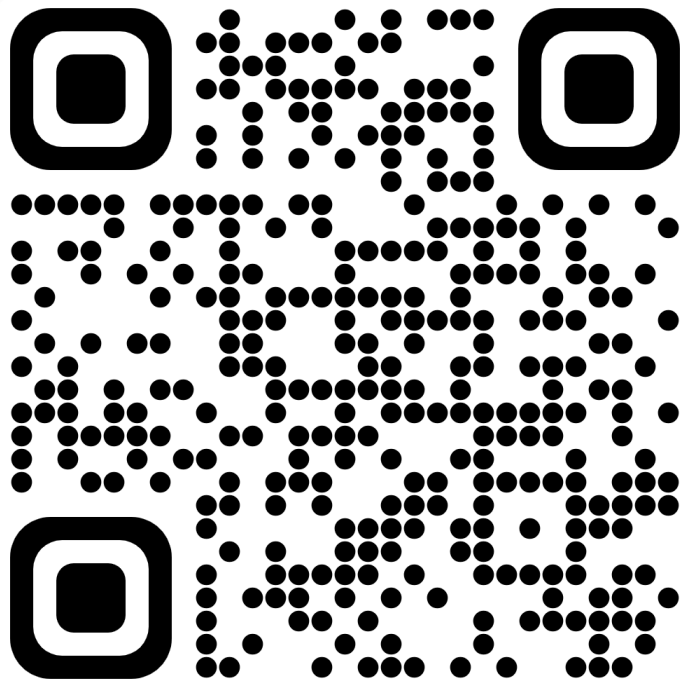
Cloud Skills Boost 的主要特點包括：

1. **豐富的學習內容：**提供各種課程，涵蓋了從初級到高級的不同技能水平，包括數據分析、機器學習、應用開發等。
2. **實踐實驗室：**學員可以在安全的虛擬環境中實踐所學，這有助於更好地理解 and 應用知識。
3. **認證準備：**平台提供了一系列的課程和資源來幫助學員準備 Google Cloud 的專業認證考試。
4. **靈活的學習路徑：**用戶可以根據自己的學習需求和速度選擇不同的課程和學習路徑。





註冊網址



<https://rsvp.withgoogle.com/events/csj-tw-s4/home>





GOOGLE FOR DEVELOPERS

Google Cloud 培訓計劃第四季：GenAI 特別篇

Google Cloud 培訓計劃 (Study Jam) 讓參加的開發人員能夠以自己的步調，學習 Google Cloud 的基礎技能。在 GenAI 特別篇中，除了學習基礎的 Google Cloud 平台之外，也增加關於 Generative AI 相關的服務。

完成指定的學習項目還能贏得 Google Cloud 相關禮品！

*請注意，30 天免費訂閱的 *access code* 將在 12 月 1 日後寄發*

📅 2023年12月1日 at 上午9:00 - 2024年1月15日 at 下午5:00 [GMT+8]

📍 TW

✓ You're registered.

[Manage your registration details](#)



為什麼您應該參加 Google Cloud 培訓計劃？

- 您可以在 Google Cloud Skills Boost 上學習 Google Cloud Platform 而不必先付費。

如何報名參加？

- 點選頁面上的「**Register**」按鈕，填寫報名表單並送出。
- 報名成功後您會收到一封報名確認電子郵件，報名確認信中含有



Google Cloud 培訓計劃第四季：
GenAI 特別篇

You're registered !

Dear 新翎,

Thank you for signing up for Google Cloud 培訓計劃第四季：GenAI 特別篇.

Now you may [click this link](#) to start the activation process of your free subscription (see the *guide below*). If you cannot click the link, please use the following URL:

https://www.cloudskillsboost.google/catalog_lab/1281?g1campaign=6p-EDUCR-GCSJ-GENAI-TWHPK-DEC2023-28

Please note:

- Please follow [this guide](#) (in Traditional Chinese) to ensure you successfully activate the 30-days free subscription.
- The Gen AI Arcade Game will open from **next week**. We will give you the access code of the arcade game later.

For any questions, you may directly contact the Qwiklabs support via support@qwiklabs.com or join our [Discord](#) #cloud-study-jam channel.

Best,

The Google Events Team

The Gen AI Arcade Game is now live! 收件匣 x



Google Cloud 培訓計劃 <no-reply-eventsatgoogle@google.com> [取消訂閱](#)

寄給我 ▾

下午4:04 (4 小時前)



Google Cloud 培訓計劃第四季：
GenAI 特別篇

Getting started!

Dear 新翊,

Thank you for joining the Google Cloud 培訓計劃第四季：GenAI 特別篇! Now you may join the Gen AI Arcade Game [via this link](#) and use the code 1q-genaius-892 to access the game. (Please note: this arcade game will only last until December 15, 2023)

For any questions, you may directly contact the Qwiklabs support via support@qwiklabs.com or join our [Discord](#) #cloud-study-jam channel.

The Google Events Team

Game

Level 3: GenAlus Chatbots

Enter an access code

This game is private. To join this game, enter an access code.

[Cancel](#)[Join](#)

first Google Cloud Gen AI credential!



Complete every activity in this game to earn a badge. Collect badges across Qwiklabs by completing quests and games. Collect them all and show off those skills!

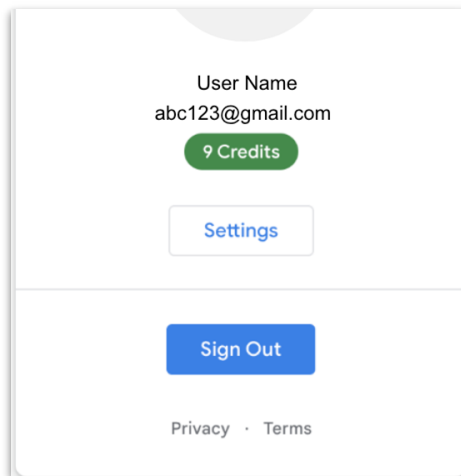
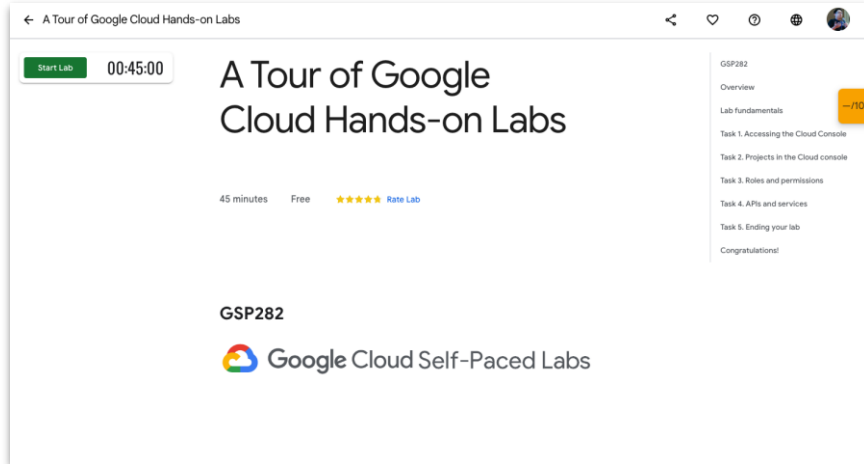
[Join this Game](#)



啟用訂閱 / Redeem Code

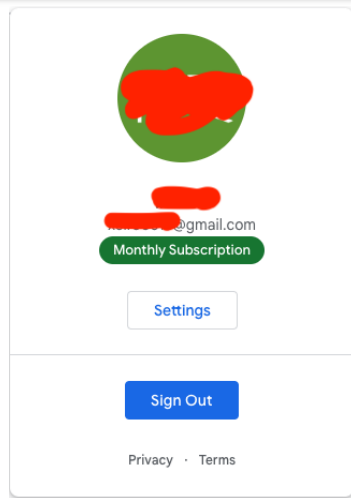
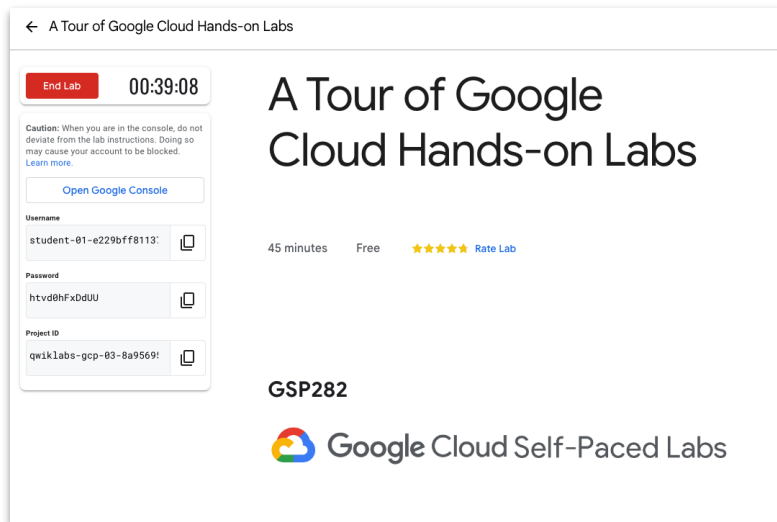
步驟 1: 點進活動給的 URL:

1. 點選活動寄發的 URL 才有 access code，進入後登入或註冊您的 Cloud Skill Boost 帳號。
2. 點擊右上角的大頭照，看看個人資訊中是否已經有 9 credits，如果沒有可以試著用無痕視窗 (Incognito window) 再試一次



步驟 2: 完成 Lab 取得一個月免費訂閱

1. 按下左上角的 **Start Lab** 的按鈕開始進行 lab。
2. 完成 lab 或是 5 分鐘後，按下左上角的 **End Lab** 按鈕，即可獲得 1 個月免費訂閱 (可在右上角的帳號 panel 確認是否有 “*Monthly Subscription*” 字樣)



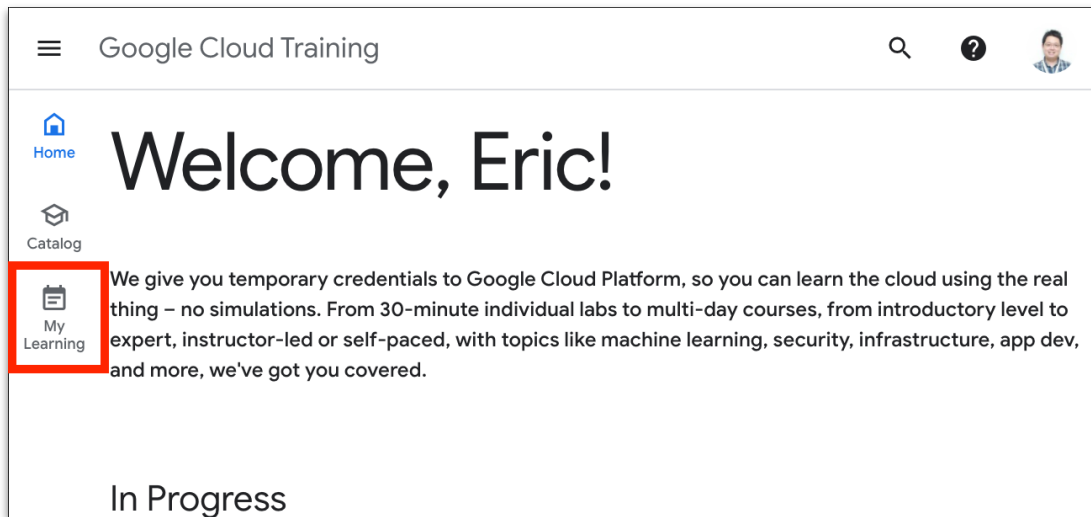
完成 Lab
5 分鐘後

檢視或回報學習成果



步驟 1: 點擊 My Learning

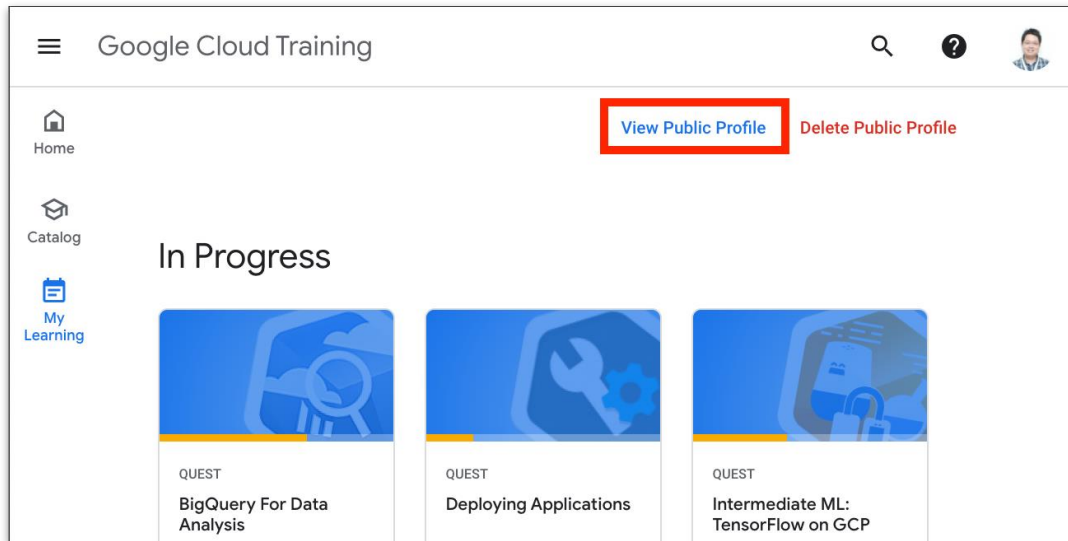
1. 登入 Qwiklabs 後，於左側選單列表中點選 **My Learning**
2. 或是直接進入
<https://google.qwiklabs.com/my-learning>



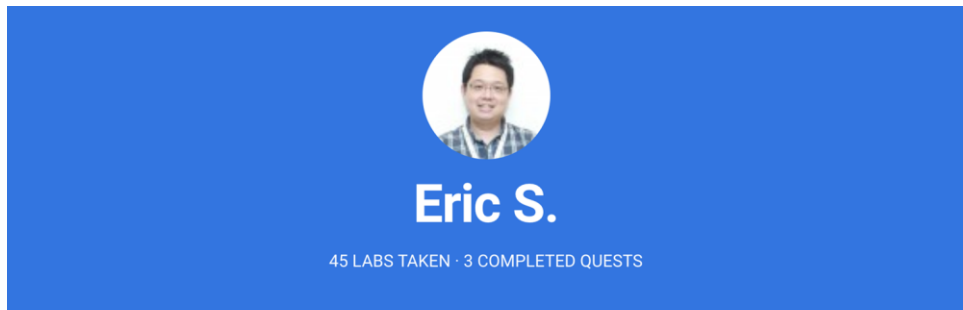
步驟 2: 點擊 View Public Profile

1. 在 My Learning 頁面中，於右上角選擇 **View Public Profile**
2. [ML Study Jam] 您可以複製這個 public profile 的 URL 回報學習成果。這個 URL 每個人都是獨一無二的，格式會像是

https://google.qwiklabs.com/public_profiles/<USER_ID>



在 **Public Profile** 中可以看到自己完成 **Quest** 所取得的 **badges**



Data Science on Google Cloud
Platform: Machine Learning
Earned Feb 14, 2019

 Add to LinkedIn



Baseline: Data, ML, AI
Earned Feb 12, 2019

 Add to LinkedIn



GCP Essentials
Earned Jan 6, 2019

 Add to LinkedIn

常見問題排解

- Q: Enroll 沒出現 Monthly Subscription
A: 先把課程 Unroll，再開無痕瀏覽模式 (Incognito Window) 再試一次。或是登出 Qwiklabs 再登入一次。
- Q: 為什麼不能 Start Lab?
A: 要先把前一次開啟的 **Lab** 關閉才能開啟新的 lab。
- Q: 自學的過程中遇到問題怎麼辦?
A: 如果是 Qwiklab 平台遇到問題，可以直接寫信給 support@qwiklabs.com 尋求協助，回應速度很快。
 - [ML Study Jam] 其它問題可以使用活動指定的討論平台或郵件群組



Skills Boost 操作小技巧





Cloud Skills Boost 操作小技巧

步驟 1: Enroll 目標學習的 Quest

1. 沒有 Enroll 的話則無法計算完成 Lab，也就不會有完成的獎勵 Badge。

Google Cloud Training

Search

Join Sign in

Home Catalog Help Center

GCP Essentials

Introductory 5 Steps 2h 50m 8 Credits

In this introductory-level quest, you will get hands-on practice with the Google Cloud Platform's fundamental tools and services. GCP Essentials is the recommended first Quest for the Google Cloud learner—you will come in with little or no prior cloud knowledge, and come out with practical experience that you can apply to your first GCP project. From writing Cloud Shell commands and deploying your first virtual machine, to running applications on Kubernetes Engine or with load balancing, GCP Essentials is a prime introduction to the platform's basic features. [1-minute videos walk you through key concepts for each lab.](#)

Infrastructure

Prerequisites

This Quest assumes little to no prior knowledge in cloud computing or with the Google Cloud Platform. It is expected that students have an information technology or computing background, and have some hands-on familiarity with administering computing systems. Prior work with shell environments / command line interfaces will be helpful for completing the labs in this series.

Enroll Now

Enroll in this quest to track your progress toward earning a badge.

[Enroll in this Quest](#)



Cloud Skills Boost 操作小技巧

步驟 2: 點選一個想要學習的 Hands-on Lab

☰ Google Cloud Training

🔍 Search



GCP Essentials

In this first hands-on lab you will access Qwiklabs and the Google Cloud Platform Console to learn about basic GCP features: Projects, Resources, IAM Users, Roles, Permissions, APIs, and Cloud Storage.



45m

Introductory

Free



HANDS-ON LAB



[Creating a Virtual Machine](#)

In this hands-on lab, you'll learn how to create a Google Compute Engine virtual machine. You'll also learn how to understand zones, regions, and machine types. To preview, watch the short video [Creating a Virtual Machine, GCP Essentials](#).



40m

Introductory

1 Credit



OR

HANDS-ON LAB

[Compute Engine: Qwik Start - Windows](#)

Google Compute Engine lets you create and run virtual machines on Google infrastructure. In this lab, you create a Windows Server instance in the Google Compute Engine and access it via a remote desktop. To preview, watch the short video [Launch a Windows Server Instance, GCP Essentials](#).



40m

Introductory

1 Credit





Cloud Skills Boost 操作小技巧

步驟 3: 啟動學習環境

1. 點擊上方的 Start Lab 按鈕啟動學習環境
2. 若在 Study Jam 贈與的 subscription 有效期間，使用 “Use Subscription” 來啟動 Lab
3. 完成 Lab 後，必須點擊上方的 End Lab 才會計入學習成果。

×

This lab costs 1 Credit.

You have a valid subscription package. Would you like to charge this lab to your subscription?

Enter Lab Access Code:

Use Subscription

Launch with Access Code



Cloud Skills Boost 操作小技巧

步驟 4: 使用無痕視窗開啟 Google Cloud Console

1. 啟動 Lab 後，平台會產生一組帳號密碼供你學習使用。
2. 建議使用**無痕視窗**開啟，並使用平台產生的帳號密碼來登入。
1. 登入後，便可照著 Cloud Skills Boost 頁面上的教學內容來學習及實驗。

Connection Details

Open Google Console

Username

google1623327_student@qwiklabs.ne

Password

TT9t5WJsCff

GCP Project ID

qwiklabs-gcp-44776a13dea667a6



Cloud Skills Boost 常見問題 (FAQ)

根據過去的經驗，如果操作不順利時，通常可以先確定以下的狀況：

1. 啟動 Lab 時要求輸入 access code 或 credits，可能是在輸入 voucher 時異常，請重新走一次啟用流程，然後確認帳號已有 **monthly subscription** 才可行。
2. 同一個帳號同一時間只能啟動一個 lab，如果前一個 lab 尚未結束、也沒有手動按 End Lab 按鈕結束，則無法啟動其它的 lab。
3. 在執行 lab 時常發生資源消失或是帳號異常的訊息，這多半是 Cloud Skills Boost 給的測試帳號與你自己本身的 Google 帳號在同一個視窗下切換錯亂，所以一般建議：
 - a. 用自己的 Google 帳號開 Lab 說明的頁面（包含啟動 lab 按鈕及有帳密的那頁）
 - b. 用**無痕視窗**（incognito window）開啟 Google Cloud Console 的環境來做 lab
4. 如果遇到的是 Qwiklab 平台或系統的問題，請直接寫信詢求協助

後續動作: 如何贏得禮物





獎勵機制

指定學習教材

GenAI Arcade Game (稍後更新)

8 個 labs · 完成時間約 2 小時

在 GenAI Arcade Game 中，你會透過一組設計好的內容學習 Google Cloud 上關於 Generative AI 的工具與服務。

[Google Cloud Computing Foundation](#)

40 個 labs · 完成時間約 30 小時

在這份教材中，你會完整學習到 Google Cloud 平台上基礎的服務，包括

- 雲端平台基礎
- Google Cloud 基礎平台服務
- Google Cloud 網路及安全服務
- Google Cloud 資料及機器學習服務


學習禮品取得資格

完成指定學習教材，根據完成的程度來給予贈品。**請注意：**禮品的寄送僅限居住在台灣以及香港的參加者。

- 您必須註冊報名這個培訓計劃、並且透過計劃給予的免費訂閱來學習
- Tier 1 - 在活動期間完成 Gen AI Arcade Game 中的學習教材，您可以獲得：
 - Google Cloud 貼紙乙份。
- Tier 2 - 在活動期間完成 Gen AI Arcade Game 以及 Google Cloud Computing Foundation 的 Unit 1 to Unit 4 教材，您可以獲得：
 - Google Cloud 貼紙乙份。
 - Google 不鏽鋼杯乙組。
- Tier 3 - 在活動期間完成 Gen AI Arcade Game 以及 Google Cloud Computing Foundation 的 Unit 1 to Unit 8 教材，您可以獲得：
 - Google Cloud 貼紙乙份。
 - Google 不鏽鋼杯乙組。
 - Google Cloud 背袋乙組。
- 透過回報表單 (稍後更新) 回報您的 Google Cloud Skill Boost 平台學習記錄，以便我們審核獲得禮品的資格。



實作: Detect Labels, Faces, and Landmarks in Images with the Cloud Vision API



實作: Vertex AI PaLM API: Qwik Start



<https://reurl.cc/DoRMod>

