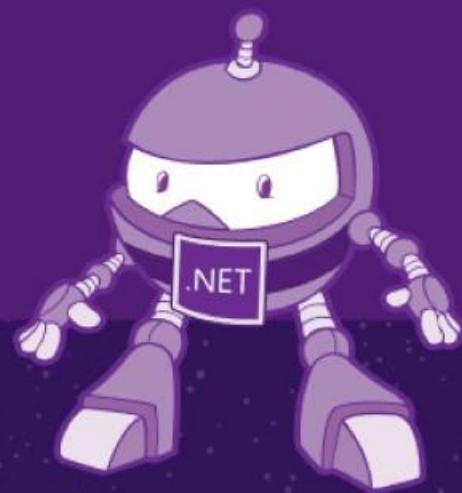


# 使用 Azure 進行文字分析與處理

歐志信



# 關於我



- 電信業工程師
- 微軟最有價值專家
  - 2019 年，獎項 Developer Technologies
  - 2018 年，獎項 Developer Technologies
  - 2017 年，獎項 Visual Studio and Development Technologies
  - 2016 年，獎項 Windows Development
  - 2015 年，獎項 Windows Platform Development
  - 2014 年，獎項 Dynamics NAV
  - 2013 年，獎項 Windows Phone Consumer
  - 2012 年，獎項 Office 365
  - 2011 年，獎項 Office System
  - 2010 年，獎項 Windows Desktop Experience
  - 2009 年，獎項 Visual C#

認知服務

Cognitive Services





## 安全性與管理



## 平台服務

### 媒體與 CDN



### 整合



### 計算服務



### 應用程式平台



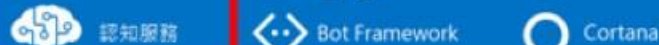
### 開發人員服務



### 資料



### 智慧



### 分析與 IoT



## 混合式雲端



## 基礎結構服務

### 計算



### 儲存體



### 網路功能



## 資料中心基礎結構



# 認知服務

- 認知服務是一套由機器學習服務支援的服務和 API，可讓開發人員將智慧型功能，例如相片和影片中的臉部辨識、文字中的情感分析，以及語言理解，納入到其應用程式

## 使用 AI 來解決業務問題



### 決策

建置應用程式，以顯示有助於做出明智與高效決策的建議。



### 辨識

辨識、識別、仲裁您的圖片、影片和數位筆跡內容，並為其製作標題及編製索引。



### 語音

將語音轉換成文字，並將文字轉換成自然發音語音。從一種語言翻譯成另一種語言，並啟用說話者驗證和辨識。



### 搜尋

將 Bing 搜尋 API 新增至您的應用程式，並充分利用以單一 API 呼叫合併數十億個網頁、影像、影片和新聞的能力。



### 語言

讓您的應用程式能夠透過預先建置的指令碼處理自然語言、評估情感，以及了解如何辨識使用者想要的項目。

<https://azure.microsoft.com/zh-tw/services/cognitive-services/>



# 為什麼選擇 Azure 認知服務?

## 簡單

透過常見的 REST API  
只需要幾行程式碼



## 靈活

支援多種程式語言與平台  
在廣泛的服務中找到適合你的  
應用的服務



## 經過考驗

由來自 Microsoft Research  
、Bing 和 Azure 機器學習等  
領域的專家建構

出色的文件、範例程式碼和社  
區支持



# 開發環境與必要條件

- 使用認知服務前，需先申請金鑰
- 使用用戶端 SDK
  - 使用用戶端程式庫 NuGet 下載認知服務套件
- 使用 REST API

# Text Analytics API





# 文字分析

## 情緒分析

此 API 傳回 0 到 1 之間的數值分數。接近 1 的分數表示正面情感；接近 0 的分數表示負面情感。

## 關鍵字詞擷取

傳回字串清單，表示輸入文字中的重要討論話題。

## 具名實體辨識

偵測文字中所有的具名實體，例如組織、人物和位置等。實體連結透過將文字關聯到網路上的其他資訊，來釐清不同的實體。

## 語言偵測

傳回偵測到的語言，共支援 120 種語言。

# 情感分析 V3 公開預覽

- 技術預覽階段的 Azure Cognitive Service 文字分析 API V3 已經開始支援正體中文，能夠辨識文章正面或負面情緒

<https://docs.microsoft.com/zh-tw/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-sentiment-analysis>

## 情感分析 V3 公開預覽

下一版情感分析現在已有公開預覽版。其大幅改善了 API 文字分類和評分的精確度和詳細資料。請試著使用 [API 測試主控台](#) 來執行。

### ⓘ 注意

- 情感分析 v3 的要求格式和**資料限制**與前版相同。
- 此外，情感分析 V3：
  - 目前支援英文(en)、日文(ja)、簡體中文(zh-Hans)、**繁體中文(zh-Hant)**、法文(fr)、義大利文(it)、西班牙文(es)、荷蘭文(nl)、葡萄牙文(pt)和德文(de)語言。
  - 適用區域如下：Australia East、Central Canada、Central US、East Asia、East US、East US 2、North Europe、Southeast Asia、South Central US、UK South、West Europe 和 West US 2。

功能	說明
更高的精確度	在文字文件中偵測正面、中性、負面和混合情感的能力顯著優於舊版。
文件和句子層級的情感分數	同時偵測文件及其個別句子的情感。如果文件包含多個句子，則會為每個句子指派情感分數。
情感標記和評分	現在，API 除了會傳回情感分數以外，還會傳回文字的情感分類。這些類別包括 positive、negative、neutral 和 mixed。
改善的輸出	現在，情感分析會同時傳回整份文字文件及其個別句子的資訊。



# 文字分析測試

## 從您的文字擷取資訊

使用下列示範來試驗 Text Analytics API。選擇其中一個範例或提供 您自己的範例。按一下 [分析] 就能識別您文字的語言、情感、關鍵片語與實體 (預覽)。

觀看影片

I had a wonderful trip to Seattle last week and even visited the Space Needle 2 times!

分析

範例 - 英文 - 正值

範例 - 英文 - 負值

範例 - 西班牙文 - 正值

範例 - 西班牙文 - 負值

分析的文字

JSON

- 語言: English (信心 : 100%)
- 關鍵片語: week, Space Needle, wonderful trip, Seattle, times
- 人氣: 84 %
- 具名實體: Seattle [Location]  
last week [DateTime-DateRange]  
Space Needle [Location]  
Space Needle [Organization]  
2 [Quantity-Number]
- 連結實體: I had a wonderful trip to Seattle last week and even

<https://azure.microsoft.com/zh-tw/services/cognitive-services/text-analytics/>



# 申請服務

首頁 > 新增 > Marketplace > 文字分析

## 文字分析

Microsoft



文字分析 [♥ 儲存以便稍後使用](#)

Microsoft

[建立](#)

[概觀](#) [方案](#)

使用簡行程式碼，輕鬆分析人聲、擷取關鍵字、識別實體及偵測任何一種文字的語言。

**情感分析**

了解使用者對您產品或服務的想法，您可為文字的整體人聲評分。

**關鍵字擷取**

自動從文字擷取關鍵字，快速找出重點。

**語言偵測**

從 120 種支援的語言中，判斷一段文字是以哪種語言寫成。

**具名實體辨識**

識別實體，並分類為人員、位置、組織、日期/時間、貨幣等，也能將已知實體，並將其連結至 Web 上可用的資訊。

**實用的連結**

[深入了解文字分析 API](#)

[文件](#)

[程式碼範例](#)

[API 參考](#)

[最新內容](#)

[定價](#)

[區域可用性](#)

Microsoft Azure

首頁 > 新增 > Marketplace > 文字分析 > Create

### Create

Text Analytics

Name \*

ta0108

訂用帳戶 \*

Visual Studio Enterprise

位置 \*

(亞太地區) 東亞

Pricing tier ([View full pricing details](#)) \*

F0 (\$K Transactions per 30 days)

Resource group \*

dtc0108

[新建](#)

[Create](#) [Automation options](#)

# 文字分析 - 程式實作

文字分析

天氣真好，我想跟 Mary 開車去故宮博物院看展覽，順便去吃小籠包

文字分析

情感分析 Sentiment:  
Sentiment Score: 0.54  
語言偵測 Detect Language:  
Language: Chinese\_Traditional  
實體辨識 Entities:

Name: Mary,	Type: Person,	Sub-Type: N/A	Offset: 9, Length: 4,	Score: 0.979
Name: 故宮博物院,	Type: Location,	Sub-Type: N/A	Offset: 17, Length: 5,	Score: 0.928

關鍵片語擷取 Key phrases:  
順便去吃小籠包 天氣真好 我想跟 Mary 開車去故宮博物院看展覽

# /Docs

## Azure 認知服務 文字分析 API 入門

[aka.ms/TextCogSvc](https://aka.ms/TextCogSvc)

 Microsoft Azure

聯絡銷售員: 1-800-857-1389

登入

概觀 解決方案 產品 文庫 定價 訓練 Marketplace 合作夥伴 支援 部落格 其他

免費

Azure / 認知服務 / 文字分析

意見反應 編輯 分享 背景主題 繁體中文 chihhsinou

依自己的步調學習 Azure 查看訓練課程模組

依標題篩選

文字分析 API 文件

概觀

什麼是文字分析 API ?

使用者案例範例

新功能

快速入門

> 用戶端程式庫

> REST API

> 教學課程

> 使用說明指南

> 參考

> 資源

下載 PDF

什麼是文字分析 API ?

2019/11/07

文字分析 API 是一種雲端式服務，可針對原始的文字提供進階的自然語言處理，同時包含四種主要功能：情感分析、關鍵片語擷取、語言偵測和實體辨識。

此 API 屬於 [Azure 認知服務](#) 的一部分，由雲端中的機器學習和 AI 演算法的集合所組成，可供您的開發專案使用。



此頁面有所助益嗎？

是 否

本文內容

情感分析

關鍵片語擷取

語言偵測

具名實體辨識

使用套餐

一般工作流程

適用於多個程式設計經驗層級的 Text Analytics

支援的語言

資料限制

Unicode 編碼



# /Learn

## 使用文字分析 API 來探索文字中的情感

<https://docs.microsoft.com/zh-tw/learn/modules/classify-user-feedback-with-the-text-analytics-api/>



900 XP

### 使用文字分析 API 來探索文字中的情感

48 分鐘 • 模組 • 8 單位

★★★★☆ 4.6 (442) 評分

中階 開發人員 AI 工程師 Azure Cosmos DB Cloud Shell Functions Azure Portal Storage Cognitive Services

了解您的客戶在傳送意見反應時，對於您產品或品牌的真正看法是什麼。我們會建立解決方案，使用 Azure Functions 和文字分析 API 的智慧功能來探索文字簡訊中的情感。

在此課程模組中，您將會：

- 了解文字分析 API
- 註冊文字分析 API 金鑰
- 使用 Azure Functions 來設計及建置服務，該服務會使用文字分析 API 來排序文字意見反應
- 透過繫結的協助，在函數應用程式中讀取及寫入 Azure 佇列儲存體訊息

必要條件

無

此單元為這些學習途徑的一部分

使用 Azure Cognitive Language Services 評估文字

簡介

2 分鐘

情感分析和文字分析 API

4 分鐘

練習 - 從線上測試主控台呼叫文字分析 API

10 分鐘

設計意見反應排序程式

6 分鐘

# OCR (光學字元辨識) API



# OCR API

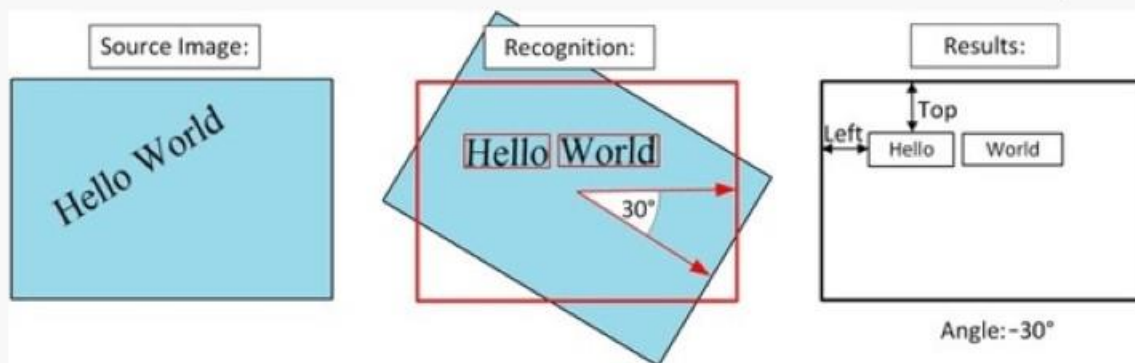
- 使用光學字元辨識 (OCR) 來偵測影像中的文字，然後將辨識出的字詞擷取到電腦可讀取的字元資料流中
- 分析影像，進而偵測內嵌的文字、產生字元資料流並啟用搜尋。替文字照相，省下複製文字所需的時間及精力
- API 網址：  
<https://westcentralus.dev.cognitive.microsoft.com/docs/services/5adf991815e1060e6355ad44/operations/56f91f2e778daf14a499e1fc>
- 中文說明：  
<https://docs.microsoft.com/zh-tw/azure/search/cognitive-search-skill-ocr>



# OCR API

## • 文字旋轉

- 若有必要，OCR 會以角度為單位傳回以水平影像座標軸為依據的旋轉位移，來修正已辨識文字的旋轉。OCR 也會提供每個字的框架座標，如下圖所示。



## • OCR API 使用須符合下列需求的影像:

- 影像必須以 JPEG、PNG、GIF 或 BMP 格式呈現。
- 輸入影像的大小必須介於 50 x 50 與 4200 x 4200 圖元之間。
- 影像中的文字能以 90 度的任何倍數進行旋轉，並可輔以不超過 40 度的小角度旋轉。

# 申請服務


Microsoft Azure

搜尋資源、服務及文件 (G+/)

首頁 > Marketplace > 電腦視覺

電腦視覺

Microsoft



電腦視覺

Microsoft

儲存以便稍後使用

建立

概觀

方案

您是否需要一個 API，能夠應用程式中使用的影像提供可操作的資訊？電腦視覺 API 提供您能理解任何影像內容的工具，建立能識別影像上物體，並清楚句子加以描述。

**使用電腦視覺 API:**

- 產生標記，以及清楚的整句影像說明。
- 讀取影像中的印刷文字
- 讀取影像中的手寫文字
- 辨識名人和地標
- 近乎即時地分析影片
- 產生縮圖

**實用的連結**

- [深入了解電腦視覺 API](#)
- [文件](#)
- [SDK](#)
- [API 參考](#)
- [定價](#)
- [區域可用性](#)

portal.azure.com/#create/Microsoft.CognitiveS

Microsoft Azure

首頁 > Marketplace > 電腦視覺 > Create

Create

Computer Vision

Name \*

dtc0108

訂用帳戶 \*

Visual Studio Enterprise

位置 \*

(美國) 美國西部 2

Pricing tier (View full pricing details) \*

F0 (20 Calls per minute, 5K Calls per month)

Resource group \*

(新增) dtc0108

新建

# OCR 實作





# Translator Text



# Translator Text

- 透過 Translator Text 開發人員很容易就能將其整合到自己的應用程式網站、工具，或是任何需要多語支援的解決方案。
- 自動偵測語言
  - 輕鬆及準確地偵測任何文字字串的語言、簡化開發程序，並可讓您快速傳送要翻譯的內容或提供當地語系化的內容。
- 轉換成不同的音義符號
  - 以不同的音義符號顯示文字，以利閱讀，像是將中文字元翻譯成拼音、以拉丁字母顯示任何支援的轉換語言，甚至以非拉丁字元 (例如日文、印度文或阿拉伯文) 顯示以拉丁文書寫的文字。

# Translator Text

- 自動偵測語言

- 輕鬆及準確地偵測任何文字字串的語言、簡化開發程序，並可讓您快速傳送要翻譯的內容或提供當地語系化的內容。

- 轉換成不同的音義符號

- 以不同的音義符號顯示文字，以利閱讀，像是將中文字元翻譯成拼音、以拉丁字母顯示任何支援的轉換語言，甚至以非拉丁字元 (例如日文、印度文或阿拉伯文) 顯示以拉丁文書寫的文字。

- 翻譯

- 將文字內容翻譯成指定語言。



# 申請服務

## Translator Text (翻譯文字)

Microsoft



### Translator Text (翻譯文字) [儲存以便稍後使用](#)

Microsoft

[建立](#)

[概觀](#) [方案](#)

Microsoft Translator API 是一項符合 ISO 與 HIPAA 規範的類神經機器翻譯 (NMT) 服務。開發人員可輕鬆將其整合至應用程式、網站、工具或是任何需要多語言支援的解決方案，例如公司網站、電子商務網站、客戶支援應用程式、傳訊應用程式、內部通訊等。

#### 提升應用程式的觸及範圍

透過 Translator API 的開放式 REST 介面，在 60 多種支援的語言間翻譯文字。

#### 轉寫成其他字母

以不同字母顯示文字使內容更易閱讀，可將漢字轉為拼音，以拉丁字母來顯示所有支援的 [韓文語言](#)，甚至可將用拉丁字母寫成的文字以非拉丁文字元顯示出來，例如阿拉伯文、印度文或日文。

#### 雙語字典

為從英文翻譯或翻譯成英文的文字找到別種翻譯，並藉由提供人類翻譯的句子範例，來找到別種文字翻譯內容。

#### 將連線與離線的類神經機器翻譯新增至您的 Android 應用程式

利用 Android 上 Microsoft Translator 應用程式的 [本機功能 \(預覽\)](#) 輕鬆將連線與離線類神經翻譯新增至您的應用程式，呼叫 Translator 應用程式將文字翻譯功能整合至您的應用程式。Translator 應用程式會在裝置連線時使用 Translator 雲端服務傳回翻譯，若離線則會透過離線套件傳回，其會使用您現有的 Translator Text API 訂閱帳戶進行收費。

#### 自動偵測語言

輕鬆且精準地偵測任何文字字串的語言，並決定字串是否需要翻譯，以及要翻成何種語言。

#### 打造自訂翻譯系統

透過 [Custom Translator](#)，您可使用現有的人類翻譯內容來打造自訂類神經系統，使寫作風格、企業形象和詞彙的處理成效更為優異。

#### 實用的連結

[深入了解 Translator Text API](#)

[文件](#)

[API 參考](#)

[定價](#)

[區域可用性](#)

# 可轉換與翻譯的語言

<div>← → ↺ docs.microsoft.com/zh-tw/azure/cognitive-services/translator/language-support</div> <div>🔍 依標題篩選</div> <div>翻譯工具文字文件</div> <div>▼ Translator Text API</div> <div>    &gt; 概觀</div> <div>    &gt; 快速入門</div> <div>    &gt; 教學課程</div> <div>    &gt; 概念</div> <div>    &gt; 使用說明指南</div> <div>    &gt; 範例</div> <div>    &gt; 參考</div> <div>▼ 資源</div> <div>    GitHub 上的程式碼範例</div> <div>    常見問題集</div> <div>    語言與區域支援</div> <div>    價格</div> <div>    區域可用性</div> <div>    法規遵循</div> <div>    要求限制</div> <div>    堆疊溢位</div> <div>    自訂翻譯工具</div>				
<h2>轉換</h2> <p>轉換方法支援以下語言。在「至/自」列中，「&lt;--&gt;」表示該語言後方所列的兩種文字可相互轉換。「--&gt;」表示該語言只能從某種文字轉寫成另一種文字。</p>				
語言	語言代碼	指令碼	至/自	指令碼
阿拉伯文	ar	阿拉伯文 Arab	<-->	拉丁文 Latn
孟加拉文	bn	孟加拉文 Beng	<-->	拉丁文 Latn
中文(簡體)	zh-Hans	簡體中文 Hans	<-->	拉丁文 Latn
中文(簡體)	zh-Hans	簡體中文 Hans	<-->	繁體中文 Hant
中文(繁體)	zh-Hant	繁體中文 Hant	<-->	拉丁文 Latn
中文(繁體)	zh-Hant	繁體中文 Hant	<-->	簡體中文 Hans
古吉拉特文	gu	古吉拉特文 Gujr	-->	拉丁文 Latn
希伯來文	he	希伯來文 Hebr	<-->	拉丁文 Latn
北印度文	hi	梵文字母 Deva	<-->	拉丁文 Latn
日文	ja	日文 Jpan	<-->	拉丁文 Latn
坎那達文	kn	坎那達文 Knda	-->	拉丁文 Latn
馬來亞拉姆文	ml	馬來亞拉姆文 Mlym	-->	拉丁文 Latn
馬拉地文	mr	梵文字母 Deva	-->	拉丁文 Latn
歐利亞文	or	歐利亞文 Orya	<-->	拉丁文 Latn

<https://docs.microsoft.com/en-us/azure/cognitive-services/translator/language-support>

# Translator Text 實作

Translator Text

天氣真好，我想開車去故宮博物院看展覽，順便去吃小籠包

偵測語言: zh-Hant  
信賴值: 1

翻譯語言 zh-Hans:  
翻譯結果: 天气真好，我想开车去故宫博物院看展览，顺便去吃小笼包

翻譯語言 ja:  
翻譯結果: 天気が良かったので、私は展覧会を見るために宮殿博物館に運転し、小さなケージバッグを食べるために立ち寄りたい

翻譯語言 en:  
翻譯結果: The weather is very good, I want to drive to the Palace Museum to see the exhibition, by the way to eat small cage bag

轉換成簡體中文

翻譯



# 内容仲裁 Content Moderator



# Content Moderator

- Content Moderator 是一項認知服務，可檢查文字、影像和視訊內容是否含有可能具冒犯意味、有風險或不當的資料。

API 群組	說明
<a href="#">文字仲裁</a>	掃描文字中具冒犯性的內容、明顯色情或性暗示內容、粗話和個人資料。
<a href="#">自訂字詞清單</a>	根據自訂字詞清單以及使用內建字詞來掃描文字。使用自訂清單可根據您自己的內容原則來封鎖或允許內容。
<a href="#">影像仲裁</a>	掃描影像中的成人或猥褻內容、使用光學字元辨識 (OCR) 功能偵測影像中的文字，以及偵測臉部。
<a href="#">自訂影像清單</a>	根據自訂影像清單掃描影像。使用自訂影像清單，篩選掉重複出現、而您不想再次分類的內容執行個體。
<a href="#">影片仲裁</a>	掃描影片中的成人或猥褻內容，並傳回該內容的時間標記。

# 文字仲裁

- 掃描文字中具冒犯性的內容、明顯色情或性暗示內容、粗話和個人資料。
- 服務回應會包含下列資訊：
  - 粗話：搭配內建的多語言粗話字詞清單進行字詞型比對
  - 分類：由電腦輔助分類來分成三種類別
  - 個人資料
  - 自動校正的文字
  - 原始文字
  - 語言

<https://docs.microsoft.com/zh-tw/azure/cognitive-services/content-moderator/text-moderation-api>



# 申請服務

Content Moderator

Microsoft



Content Moderator

Microsoft

儲存以便稍後使用

建立

概觀

方案

影像仲裁

透過以機器學習服務為基礎的分類器、自訂封鎖清單及光學字元辨識 (OCR)，讓您更能夠偵測到可能有冒犯意味或不應出現的影像。

文字仲裁

協助您偵測超過 100 種語言中潛在的不雅內容，並自動對您的自訂清單進行比對。內容仲裁者也會檢查是否可能有個人識別資訊 (PII)。

影片仲裁

能夠對影片中可能出現的成人內容評分，影片仲裁目前在 Azure 媒體服務以預覽形式部署。

審核工具

在預測信賴度須經過協助來證明的情況下，或在必須依現實情境放寬決定時，人力可以有效補強機器學習模型，透過監督演算法結果的人力審核工具，您可享有可見度、彈性和控制權。

實用的連結

[深入了解 Content Moderator](#)

[文件](#)

[API 參考](#)

[定價](#)

[區域可用性](#)

# 文字仲裁 - 程式實作



<https://docs.microsoft.com/zh-tw/azure/cognitive-services/content-moderator/dotnet-sdk-quickstart>

# 學習資源



# /Docs

## 入門系列...

- Azure 認知服務文字分析 API: <https://aka.ms/TextCogSvc>
- 自動機器學習: <https://aka.ms/AutomatedMLDoc>
- Azure 機器學習服務和 Python SDK: <https://aka.ms/AA3dzht>

## 其他資料

### Azure 機器學習最佳做法存儲庫

- 自然語言 [github.com/microsoft/nlp-recipes/](https://github.com/microsoft/nlp-recipes/)
- 電腦視覺 [github.com/microsoft/ComputerVision](https://github.com/microsoft/ComputerVision)
- 關於機器學習的 9 個進階提示: <https://aka.ms/9MLTips>



# /MS Learn

完成互動式學習練習，觀看影片，  
並練習和運用學到的新技能。

[aka.ms/AIML40MSLearnCollection](https://aka.ms/AIML40MSLearnCollection)

## 此集合中的課程模組



### Azure Machine Learning 服務簡介

模組 • 35 分鐘

★★★★☆ 4.6 (744)

0% 已完成



### 使用 Azure Machine Learning 服務將本機 ML 模型定型

模組 • 53 分鐘

★★★★☆ 4.6 (461)

0% 已完成



### 使用 Azure Machine Learning 服務註冊及部署 ML 模型

模組 • 1 小時 24 分鐘

★★★★☆ 4.5 (388)

0% 已完成



### 使用 Azure Machine Learning 服務自動選取 ML 模型

模組 • 25 分鐘

★★★★☆ 4.6 (390)

# 延伸學習

- Azure 認知搜尋

- <https://azure.microsoft.com/zh-tw/services/search/>
- <https://github.com/microsoft/ignite-learning-paths-training-aiml/blob/master/aiml10/demo1.md>

- 表單辨識器 預覽

- <https://azure.microsoft.com/zh-tw/services/cognitive-services/form-recognizer/>
- <https://github.com/microsoft/ignite-learning-paths-training-aiml/blob/master/aiml10/demo2.md>



# Microsoft Azure

---

Invent with purpose.