



將 Amazon Comprehend Medical 和 LLMs用於醫療保健和生命科學

# AWS 方案指引



# AWS 方案指引: 將 Amazon Comprehend Medical 和 LLMs用於醫療保健和生命科學

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商標和商業外觀不得用於任何非 Amazon 的產品或服務，也不能以任何可能造成客戶混淆、任何貶低或使 Amazon 名譽受損的方式使用 Amazon 的商標和商業外觀。所有其他非 Amazon 擁有的商標均為其各自擁有者的財產，這些擁有者可能附屬於 Amazon，或與 Amazon 有合作關係，亦或受到 Amazon 贊助。

# Table of Contents

簡介 .....	1
概觀 .....	1
目標對象 .....	1
目標 .....	2
技術方法 .....	3
使用 Amazon Comprehend Medical .....	3
功能 .....	4
使用案例 .....	5
結合 Amazon Comprehend Medical 與 LLMs .....	5
Architecture .....	6
使用案例 .....	7
最佳 - 實務 .....	7
提示 - 工程 .....	8
使用 LLMs .....	17
LLM 的使用案例 .....	17
自訂 .....	17
選擇 LLM .....	19
微調 LLMs .....	22
估算成本和投資報酬率 .....	22
選擇策略 .....	23
建置資料集 .....	24
微調 .....	25
監控 .....	26
選擇合適方案 .....	27
業務成熟度考量事項 .....	28
評估 LLMs .....	30
訓練和測試資料 .....	30
指標 .....	31
常見問答集 .....	32
如何在 Amazon Comprehend Medical 和 LLM 之間進行選擇？ .....	32
如何將 Amazon Comprehend Medical 結果提供給 LLM？ .....	32
搭配使用 Amazon Comprehend Medical 與 LLMs 時，有哪些最佳實務？ .....	32
我是否應該使用預先訓練的醫療 LLM，或針對我的醫療保健使用案例微調一般 LLM？ .....	32
如何評估醫療 NLP 任務的 LLMs 效能？ .....	32

高複雜性和低複雜性 LLM 解決方案之間的權衡是什麼？ .....	33
後續步驟 .....	34
AWS 資源 .....	34
其他資源 .....	35
貢獻者 .....	36
編寫 .....	36
檢閱 .....	36
技術寫入 .....	36
文件歷史紀錄 .....	37
詞彙表 .....	38
# .....	38
A .....	38
B .....	41
C .....	42
D .....	45
E .....	48
F .....	50
G .....	51
H .....	52
I .....	53
L .....	55
M .....	56
O .....	60
P .....	62
Q .....	64
R .....	64
S .....	67
T .....	70
U .....	71
V .....	71
W .....	72
Z .....	73
.....	lxxiv

# 將 Amazon Comprehend Medical 和 LLMs 用於醫療保健和生命科學

Amazon Web Services [???](#) ( 貢獻者 )

2025 年 12 月 ([文件歷史記錄](#))

## 概觀

不斷增加的醫療資料量以及對有效和準確處理的需求，推動了採用人工智慧和機器學習 (AI/ML) 技術的 [自然語言處理 \(NLP\)](#)。預先訓練的分類器模型和 [大型語言模型 \(LLMs\)](#) 已成為各種醫療 NLP 任務的強大工具，包括臨床問題答案、報告摘要和洞見產生。不過，醫療保健和生命科學領域由於醫學術語、領域特定知識和法規要求的複雜性，帶來了獨特的挑戰。在此網域中有效使用預先訓練的分類器或 LLMs 需要精心設計的方法，將這些模型的優勢與特定網域的資源和技術結合在一起。

醫療保健和生命科學產業實務傳統上依賴規則型系統、手動編碼和專家審核程序。這些系統和程序耗時且容易出錯。AI 和 NLP 技術的整合，例如 [Amazon Comprehend Medical](#) 和 [Amazon Bedrock](#) 中的基礎模型，提供高效且可擴展的解決方案，可處理醫療資料，同時提高準確性和一致性。

本指南探討在醫療保健產業中使用 Amazon Comprehend Medical 和 LLMs 進行智慧型自動化。它概述了簡化醫療編碼、病患資訊擷取和記錄摘要程序的最佳實務、挑戰和實際方法。透過使用 Amazon Comprehend Medical 和 LLMs 的功能，醫療保健組織可以釋放新層級的操作效率、降低成本，並可能改善患者護理。

本指南詳細說明醫療保健領域的獨特考量，例如了解醫學術語、使用特定領域的 LLMs，以及解決 AI/ML 系統的限制。它為醫療保健 IT 經理、架構師和技術主管提供全面的決策路徑，以評估組織準備程度、評估實作選項，並使用適當的 AWS 服務 和 工具來成功自動化。

透過遵循本指南中概述的指導方針和最佳實務，醫療保健組織可以利用 AI/ML 技術的強大功能，同時導覽醫療領域的複雜性。此方法支援遵守道德和法規準則，並促進在醫療保健中負責任地使用 AI 系統。它旨在產生準確且私密的洞見。

## 目標對象

本指南適用於希望實作 AI 驅動之自然語言處理解決方案以進行醫療資料分析和自動化的技術利益相關者、架構師、技術主管和決策者。

## 目標

醫療保健和生命科學組織可以使用 Amazon Comprehend Medical 和 LLMs 來實現多個業務目標。這些結果通常包括提高營運效率、降低成本和改善患者護理。本節概述關鍵業務目標，以及實作本指南中概述之策略和最佳實務的相關優勢。

以下是組織可透過實作本指南中的指導方針和最佳實務來實現的一些目標：

- 縮短開發時間 – 本指南的最終目標是降低成本縮短開發時間、減少技術負債，以及緩解 POC 的潛在專案故障。透過了解 Amazon Comprehend Medical 等關鍵 AI/ML 服務，以及 LLM 用於醫療保健任務的優點和限制，企業可以更快地上市並提高速度，以達成業務目標。
- 擷取資訊以自動化醫療編碼任務 – 在患者就診之後，編碼專家和提供者可以從醫療文字中擷取洞見，例如主觀、目標、評估和計劃 (SOAP) 備註。這可以減少手動文件工作，並協助提供者專注於病患的需求。透過結合 Amazon Comprehend Medical 的實體辨識功能與 LLMs，組織可以從患者記錄、臨床備註和其他醫療保健資料來源中擷取相關的醫療資訊。這可將人為錯誤降至最低，並促進一致的實務。
- 摘要患者記錄和臨床文件 – 自動化摘要患者歷程記錄、治療計劃和醫療結果，可為醫療保健供應商節省寶貴的時間。LLMs 可協助產生完整且結構化的臨床文件。您可以使用 Amazon Comprehend Medical 取得其他內容、使用醫療網域 LLM，或使用醫療資料微調 LLM。這些方法有助於提供準確的摘要，並確保文件符合合規要求和標準。
- 支援臨床決策和患者護理 – 透過在 Amazon Comprehend Medical 中使用[內科連結](#)，以及使用 LLMs，提供者可以回答醫療問題或尋求解決患者護理的建議。這可讓醫療專業人員做出明智的決策，以改善患者結果並降低醫療錯誤的風險。

# 醫療保健和生命科學的生成式 AI 和 NLP 方法

自然語言處理 (NLP) 是一種機器學習技術，可讓電腦解譯、操作和理解人類語言。醫療保健和生命科學組織擁有來自患者記錄的大量資料。他們可以使用 NLP 軟體自動處理此資料。例如，他們可以將 NLP 與生成式 AI 結合，以簡化醫療編碼、擷取患者資訊和摘要記錄。

根據您要執行的 NLP 任務，不同的架構可能最適合您的使用案例。本指南針對醫療和生命科學應用程式的下列生成式 AI 和 NLP 選項 AWS：

- [使用 Amazon Comprehend Medical](#) – 了解如何獨立使用 Amazon Comprehend Medical，而不將其與大型語言模型 (LLM) 整合。
- [結合 Amazon Comprehend Medical 與大型語言模型](#) – 了解如何在擷取增強產生 (RAG) 架構中結合 Amazon Comprehend Medical 與 LLM。
- [針對醫療保健和生命科學使用案例使用大型語言模型](#) – 了解如何使用經過微調的 LLM 或 RAG 架構，將 LLM 用於醫療保健和生命科學應用程式。

## 使用 Amazon Comprehend Medical

[Amazon Comprehend Medical](#) 是一種 AWS 服務，可偵測並傳回非結構化臨床文字中的有用資訊，例如醫生備註、出院摘要、測試結果和案例備註。它使用自然語言處理 (NLP) 模型來偵測實體。實體是醫療資訊的文字參考，例如醫療狀況、藥物或受保護醫療資訊 (PHI)。

### Important

Amazon Comprehend Medical 無法取代專業醫療建議、診斷或治療。Amazon Comprehend Medical 提供可信度分數，指出偵測到之實體準確性的可信度。確認使用案例的相關可信度閾值，並在需要高準確性的情況下使用高可信度閾值。在某些情況下，應由經過適當訓練的人工檢閱者檢閱和驗證結果。例如，Amazon Comprehend Medical 只有在經過訓練的醫療專業人員審查準確性和健全的醫學判斷後，才應該用於患者護理案例。

您可以透過 AWS 管理主控台、AWS Command Line Interface (AWS CLI) 或 SDK 存取 AWS SDKs Amazon Comprehend Medical。此 AWS SDKs 適用於各種程式設計語言和平台，例如 Java、Python、Ruby、.NET、iOS 和 Android。您可以使用 SDKs 從用戶端應用程式以程式設計方式存取 Amazon Comprehend Medical。

本節會檢閱 Amazon Comprehend Medical 的主要功能。它還討論了與大型語言模型 (LLM) 相比，使用此服務的優勢。

## Amazon Comprehend Medical 功能

Amazon Comprehend Medical 提供近乎即時和批次推論 APIs。這些 APIs 可以使用醫療實體辨識和識別實體關係，擷取醫療文字並提供醫療 NLP 任務的結果。您可以對單一檔案執行分析，或對存放在 Amazon Simple Storage Service (Amazon S3) 儲存貯體中的多個檔案執行批次分析。Amazon Comprehend Medical 提供下列用於同步實體偵測的文字分析 API 操作：

- [偵測實體](#) – 偵測一般醫療類別，例如結構、醫療情況、PHI 類別、程序和時間表達式。
- [偵測 PHI](#) – 偵測特定實體，例如年齡、日期、名稱和類似的個人資訊。

Amazon Comprehend Medical 也包含多個 API 操作，可用於對臨床文件執行批次文字分析。若要進一步了解如何使用這些 API 操作，請參閱[文字分析批次 APIs](#)。

使用 Amazon Comprehend Medical 偵測臨床文字中的實體，並將這些實體連結至標準化醫療拓撲中的概念，包括 RxNorm、ICD-10-CM 和 SNOMED CT 知識庫。您可以對單一檔案執行分析，或對存放在 Amazon S3 儲存貯體中的大型文件或多個檔案執行批次分析。Amazon Comprehend Medical 提供下列本體連結 API 操作：

- [InferICD10CM](#) – InferICD10CM 操作會偵測潛在的醫療情況，並將其連結至 2019 年國際疾病分類第 10 版臨床修改 (ICD-10-CM) 的代碼。對於每個偵測到的潛在醫療情況，Amazon Comprehend Medical 會列出相符的 ICD-10-CM 代碼和描述。結果中列出的醫療情況包括可信度分數，這表示 Amazon Comprehend Medical 對結果中相符概念之實體準確性的可信度。
- [InferRxNorm](#) – InferRxNorm 操作可識別在病患記錄中列為實體的藥品。它將實體連結至來自國家醫學圖書館 RxNorm 資料庫的概念識別符 (RxCUI)。每個 RxCUI 對於不同的強度和用量表單都是唯一的。結果中列出的藥物包含可信度分數，表示 Amazon Comprehend Medical 對符合 RxNorm 知識庫概念之實體準確性的可信度。Amazon Comprehend Medical 會根據可信度分數，以遞減順序列出可能符合其偵測到之每種藥物的最上層 RxCUIs。
- [InferSNOMEDCT](#) – InferSNOMEDCT 操作會將可能的醫療概念識別為實體，並將其連結至 2021 年 3 年版本的 Systematized Nomenclature of Medicine, clinical Terms (SNOMED CT) 程式碼。SNOMED CT 提供醫療概念的完整詞彙，包括醫療條件和解剖，以及醫療測試、治療和程序。對於每個相符的概念 ID，Amazon Comprehend Medical 會傳回前五個醫療概念，每個概念都有可信度分數和情境資訊，例如特徵和屬性。然後，當與 SNOMED CT 多階層搭配使用時，SNOMED CT 概念 IDs 可用於建構患者臨床資料以進行醫療編碼、報告或臨床分析。

如需詳細資訊，請參閱 Amazon Comprehend Medical 文件中的 [文字分析 APIs](#) 和 [Ontology Linking APIs](#)。

## Amazon Comprehend Medical 的使用案例

作為獨立服務，Amazon Comprehend Medical 可能會解決組織的使用案例。Amazon Comprehend Medical 可以執行下列任務：

- 協助在病患記錄中進行醫療編碼
- 偵測受保護的健康資訊 (PHI) 資料
- 驗證藥物，包括屬性，例如用量、頻率和形式

Amazon Comprehend Medical 結果適用於大多數醫療實務。不過，如果您有以下限制，您可能需要考慮替代方案：

- 不同的實體定義 – 例如，您對藥物實體 FREQUENCY 的定義可能不同。對於頻率，Amazon Comprehend Medical 會視需要進行預測，但您的組織可能會使用 pro re nata (PRN) 一詞。
- 結果數量過大 – 例如，患者備註經常包含多個症狀和關鍵字，對應到多個 ICD-10-CM 代碼。不過，數個關鍵字不適用於診斷。在此情況下，提供者必須評估許多 ICD-10-CM 實體及其可信度分數，這需要手動處理時間。
- 自訂實體或 NLP 任務 – 例如，供應商可能想要擷取 PRN 證據，例如視需要採取止痛措施。由於 Amazon Comprehend Medical 不提供此功能，因此需要不同的 AI/ML 模型。如果 NLP 任務不在實體辨識範圍內，例如摘要、問答和情緒分析，則需要不同的 AI/ML 解決方案。

## 結合 Amazon Comprehend Medical 與大型語言模型

[NEJM AI 的 2024 年研究](#)顯示，使用 LLM 搭配零鏡頭提示，進行醫療編碼任務通常會導致效能不佳。搭配 LLM 使用 Amazon Comprehend Medical 有助於緩解這些效能問題。Amazon Comprehend Medical 結果是執行 NLP 任務之 LLM 的有用內容。例如，將 Amazon Comprehend Medical 的內容提供給大型語言模型可協助您：

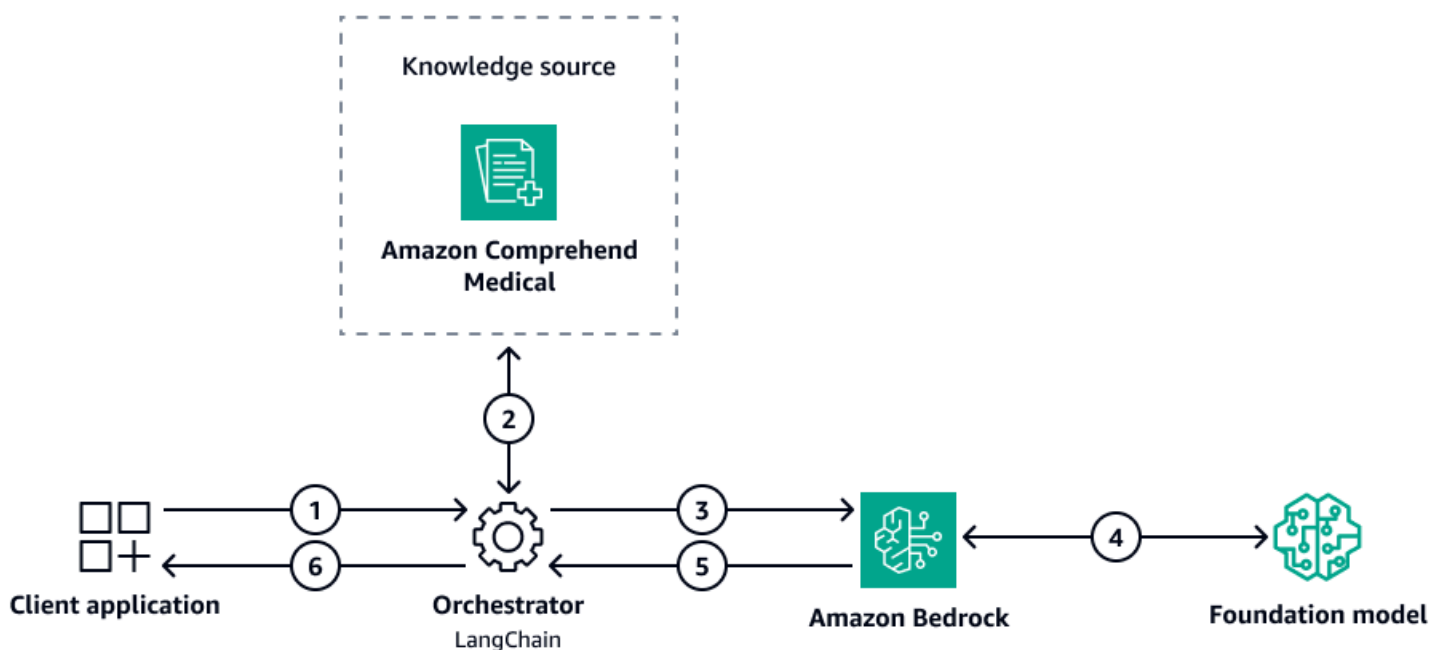
- 使用 Amazon Comprehend Medical 的初始結果作為 LLM 的內容，以提高實體選擇的準確性
- 實作自訂實體辨識、摘要、問答和其他使用案例

本節說明如何使用擷取增強生成 (RAG) 方法，將 Amazon Comprehend Medical 與 LLM 結合。擷取增強生成 (RAG) 是一種生成式 AI 技術，其中 LLM 會在產生回應之前參考訓練資料來源以外的授權資料來源。如需詳細資訊，請參閱[什麼是 RAG](#)。

為了說明此方法，本節使用與 ICD-10-CM 相關的醫療（診斷）編碼範例。它包含範例架構和提示工程範本，以協助加速您的創新。它還包括在 RAG 工作流程中使用 Amazon Comprehend Medical 的最佳實務。

## 使用 Amazon Comprehend Medical 的 RAG 架構

下圖說明從患者備註識別 ICD-10-CM 診斷碼的 RAG 方法。它使用 Amazon Comprehend Medical 作為知識來源。在 RAG 方法中，擷取方法通常會從包含適用知識的向量資料庫中擷取資訊。此架構不使用向量資料庫，而是使用 Amazon Comprehend Medical 進行擷取任務。協調器會將患者備註資訊傳送至 Amazon Comprehend Medical，並擷取 ICD-10-CM 程式碼資訊。協調器會透過 Amazon Bedrock 將此內容傳送至下游基礎模型 (LLM)。LLM 會使用 ICD-10-CM 程式碼資訊產生回應，並將該回應傳回用戶端應用程式。



圖表顯示下列 RAG 工作流程：

- 用戶端應用程式會將病患備註作為查詢傳送至協調器。這些患者備註的範例可能是「患者是 X 醫生的 71 歲女性患者。該患者昨天晚上來到了緊急室，有大約 7 天到 8 天的歷史，這是持續性的。她沒有明確的退燒或發冷，也沒有黃疸史。病患拒絕任何最近顯著的減重。」
- 協調器使用 Amazon Comprehend Medical 擷取與查詢中醫療資訊相關的 ICD-10-CM 代碼。它使用 InferICD10CM API 從患者備註中擷取和推斷 ICD-10-CM 代碼。

3. 協調器會建構提示，其中包含提示範本、原始查詢，以及從 Amazon Comprehend Medical 擷取的 ICD-10-CM 代碼。它會將此增強型內容傳送至 Amazon Bedrock。
4. Amazon Bedrock 會處理輸入，並使用基礎模型產生回應，其中包含 ICD-10-CM 代碼及其來自查詢的對應證據。產生的回應包含已識別的 ICD-10-CM 代碼，以及支援每個代碼之患者備註的證據。以下是範例回應：

```
<response>
<icd10>
<code>R10.9</code>
<evidence>history of abdominal pain</evidence>
</icd10>
<icd10>
<code>R10.30</code>
<evidence>history of abdominal pain</evidence>
</icd10>
</response>
```

5. Amazon Bedrock 會將產生的回應傳送至協調器。
6. 協調器會將回應傳回用戶端應用程式，使用者可以在其中檢閱回應。

## 在 RAG 工作流程中使用 Amazon Comprehend Medical 的使用案例

Amazon Comprehend Medical 可以執行特定的 NLP 任務。如需詳細資訊，請參閱 [Amazon Comprehend Medical 的使用案例](#)。

您可能想要將 Amazon Comprehend Medical 整合到進階使用案例的 RAG 工作流程中，如下所示：

- 將擷取的醫療實體與來自患者記錄的內容資訊結合，以產生詳細的臨床摘要
- 使用擷取的實體搭配程式碼指派的內科連結資訊，自動化複雜案例的醫療編碼
- 使用擷取的醫療實體，從非結構化文字自動建立結構化臨床備註
- 根據擷取藥物名稱和屬性來分析藥物副作用
- 開發智慧型臨床支援系統，結合擷取的醫療資訊與up-to-date研究和指導方針

## 在 RAG 工作流程中使用 Amazon Comprehend Medical 的最佳實務

將 Amazon Comprehend Medical 結果整合至 LLM 的提示時，請務必遵循最佳實務。這可以改善效能和準確性。以下是主要建議：

- 了解 Amazon Comprehend Medical 可信度分數 – Amazon Comprehend Medical 為每個偵測到的實體和本體連結提供可信度分數。請務必了解這些分數的意義，並為您的特定使用案例建立適當的閾值。可信度分數有助於篩選出低可信度實體，減少雜訊並改善 LLM 輸入的品質。
- 在提示詞工程中使用可信度分數 – 在製作 LLM 提示時，請考慮將 Amazon Comprehend Medical 可信度分數合併為其他內容。這有助於 LLM 根據實體的可信度等級排定優先順序或權重實體，從而可能改善輸出的品質。
- 使用 Ground Truth 資料評估 Amazon Comprehend Medical 結果 – Ground Truth 資料是已知為 true 的資訊。它可用來驗證 AI/ML 應用程式是否產生準確的結果。將 Amazon Comprehend Medical 結果整合到您的 LLM 工作流程之前，請在代表性的資料範例上評估服務的效能。將結果與 Ground Truth 註釋進行比較，以識別潛在的差異或需要改進的區域。此評估可協助您了解 Amazon Comprehend Medical 對您使用案例的優勢和限制。
- 從策略上選取相關資訊 – Amazon Comprehend Medical 可以提供大量資訊，但並非所有資訊都與您的任務相關。仔細選取與您的使用案例最相關的實體、屬性和中繼資料。向 LLM 提供太多不相關的資訊可能會產生雜訊，並可能降低效能。
- 對齊實體定義 – 確保 Amazon Comprehend Medical 使用的實體和屬性定義與您的解釋保持一致。如果有差異，請考慮提供其他內容或釐清給 LLM，以彌補 Amazon Comprehend Medical 輸出與您的需求之間的差距。如果 Amazon Comprehend Medical 實體不符合您的期望，您可以在提示中包含其他指示（和可能的範例），以實作自訂實體偵測。
- 提供特定領域的知識 – 雖然 Amazon Comprehend Medical 提供寶貴的醫療資訊，但可能無法擷取特定網域的所有細微差別。考慮使用額外的領域特定知識來源來補充 Amazon Comprehend Medical 結果，例如拓撲、術語或專家策劃的資料集。這可為 LLM 提供更全面的內容。
- 遵守道德和法規準則 – 處理醫療資料時，請務必遵守道德原則和法規準則，例如與資料隱私權、安全和在醫療保健中負責任使用 AI 系統相關的準則。確保您的實作符合相關法律和產業最佳實務。

透過遵循這些最佳實務，AI/ML 從業人員可以有效地使用 Amazon Comprehend Medical 和 LLMs 的優勢。對於醫療 NLP 任務，這些最佳實務有助於降低潛在風險並提高效能。

## Amazon Comprehend Medical 內容的提示詞工程

[提示詞工程](#)是設計和精簡提示詞的程序，以引導生成式 AI 解決方案產生所需的輸出。您可以選擇最適合的格式、片語、單字和符號，引導 AI 更有意義的與使用者互動。

視您執行的 API 操作而定，Amazon Comprehend Medical 會傳回偵測到的實體、本體程式碼和描述，以及可信度分數。當您的解決方案調用目標 LLM 時，這些結果會在提示中變成內容。您必須設計提示，以在提示範本中呈現內容。

**Note**

本節中的範例提示遵循 [Anthropic 指引](#)。如果您使用的是不同的 LLM 供應商，請遵循該供應商的建議。

一般而言，您會將原始醫療文字和 Amazon Comprehend Medical 結果插入提示中。以下是常見的提示結構：

```
<medical_text>
medical text
</medical_text>

<comprehend_medical_text_results>
comprehend medical text results
</comprehend_medical_text_results>

<prompt_instructions>
prompt instructions
</prompt_instructions>
```

本節提供將 Amazon Comprehend Medical 結果納入下列常見醫療 NLP 任務的提示內容的策略：

- [篩選 Amazon Comprehend Medical 結果](#)
- [使用 Amazon Comprehend Medical 擴展醫療 NLP 任務](#)
- [使用 Amazon Comprehend Medical 套用護欄](#)

## 篩選 Amazon Comprehend Medical 結果

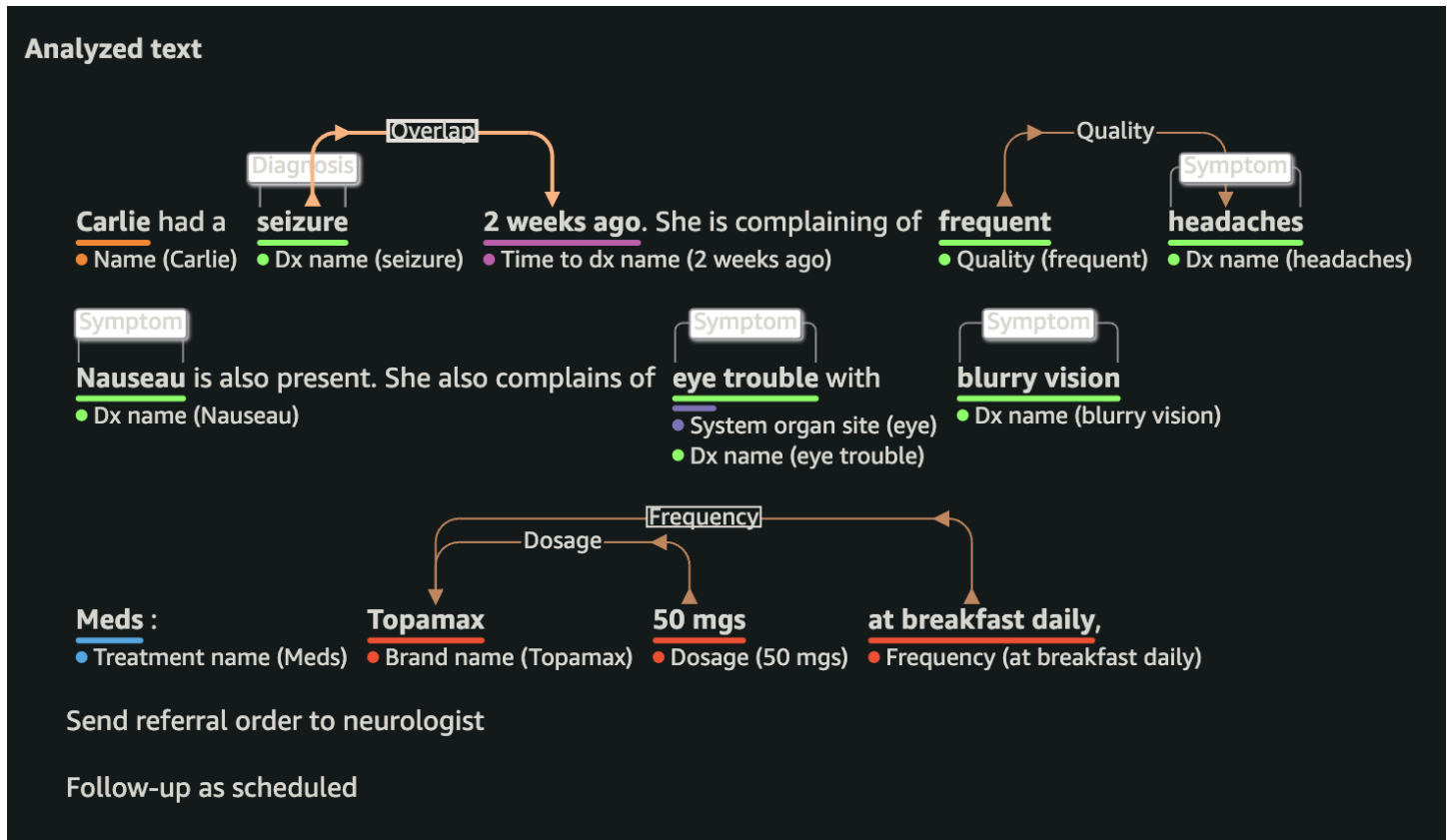
Amazon Comprehend Medical 通常會提供大量資訊。您可能想要減少醫療專業人員必須檢閱的結果數量。在這種情況下，您可以使用 LLM 來篩選這些結果。Amazon Comprehend Medical 實體包含可信度分數，您可以在設計提示時用作篩選機制。

以下是範例病患備註：

```
Carlie had a seizure 2 weeks ago. She is complaining of frequent headaches
Nausea is also present. She also complains of eye trouble with blurry vision
Meds : Topamax 50 mgs at breakfast daily,
Send referral order to neurologist
```

Follow-up as scheduled

在此患者備註中，Amazon Comprehend Medical 偵測到下列實體。



實體會連結至下列 ICD-10-CM 程式碼，以因應癱瘓和頭昏。

Category	ICD-10-CM 程式碼	ICD-10-CM 描述	可信度分數
沒收	R56.9	未指定的抽搐	0.8348
沒收	G40.909	癱瘓、未指定、不難治、沒有狀態癱瘓	0.5424
沒收	R56.00	簡單熱性抽搐	0.4937
沒收	G40.09	其他絆倒	0.4397
沒收	G40.409	其他非難治的廣義型驚悚和驚悚症狀，沒有狀態驚悚	0.4138

耳機	R51	耳機	0.4067
耳機	R51.9	耳機，未指定	0.3844
耳機	G44.52	新的每日持續性問題 (NDPH)	0.3005
耳機	G44	其他頭痛綜合徵	0.2670
耳機	G44.8	其他指定的頭痛綜合徵	0.2542

您可以將 ICD-10-CM 代碼傳遞到提示中，以提高 LLM 精確度。若要減少雜訊，您可以使用 Amazon Comprehend Medical 結果中包含的可信度分數來篩選 ICD-10-CM 代碼。以下是提示範例，只包含可信度分數高於 0.4 的 ICD-10-CM 代碼：

```
<patient_note>
Carlie had a seizure 2 weeks ago. She is complaining of frequent headaches
Nausea is also present. She also complains of eye trouble with blurry vision
Meds : Topamax 50 mgs at breakfast daily,
Send referral order to neurologist
Follow-up as scheduled
</patient_note>

<comprehend_medical_results>
<icd-10>
  <entity>
    <text>seizure</text>
    <code>
      <description>Unspecified convulsions</description>
      <code_value>R56.9</code_value>
      <score>0.8347607851028442</score>
    </code>
    <code>
      <description>Epilepsy, unspecified, not intractable, without status epilepticus</
description>
      <code_value>G40.909</code_value>
      <score>0.542376697063446</score>
    </code>
    <code>
      <description>Other seizures</description>
      <code_value>G40.89</code_value>
```

```
<score>0.43966275453567505</score>
</code>
<code>
  <description>Other generalized epilepsy and epileptic syndromes, not intractable,
without status epilepticus</description>
  <code_value>G40.409</code_value>
  <score>0.41382506489753723</score>
</code>
</entity>
<entity>
  <text>headaches</text>
  <code>
    <description>Headache</description>
    <code_value>R51</code_value>
    <score>0.4066613018512726</score>
  </code>
</entity>
<entity>
  <text>Nausea</text>
  <code>
    <description>Nausea</description>
    <code_value>R11.0</code_value>
    <score>0.6460834741592407</score>
  </code>
</entity>
<entity>
  <text>eye trouble</text>
  <code>
    <description>Unspecified disorder of eye and adnexa</description>
    <code_value>H57.9</code_value>
    <score>0.6780954599380493</score>
  </code>
  <code>
    <description>Unspecified visual disturbance</description>
    <code_value>H53.9</code_value>
    <score>0.5871203541755676</score>
  </code>
  <code>
    <description>Unspecified disorder of binocular vision</description>
    <code_value>H53.30</code_value>
    <score>0.5539672374725342</score>
  </code>
</entity>
<entity>
```

```

<text>blurry vision</text>
<code>
  <description>Other visual disturbances</description>
  <code_value>H53.8</code_value>
  <score>0.9001834392547607</score>
</code>
</entity>
</icd-10>
</comprehend_medical_results>

<prompt>
Given the patient note and Amazon Comprehend Medical ICD-10-CM code results above,
please select the most relevant ICD-10-CM diagnosis codes for the patient.
For each selected code, provide a brief explanation of why it is relevant based on the
information in the patient note.
</prompt>

```

## 使用 Amazon Comprehend Medical 擴展醫療 NLP 任務

處理醫療文字時，Amazon Comprehend Medical 的內容可協助 LLM 選取更好的字符。在此範例中，您想要將診斷症狀與藥物配對。您也想要尋找與醫療測試相關的文字，例如與血小板測試相關的術語。您可以使用 Amazon Comprehend Medical 來偵測實體和藥物名稱。在此情況下，您會使用適用於 Amazon Comprehend Medical 的 [DetectEntitiesV2](#) 和 [InferRxNorm](#) APIs。

以下是範例病患備註：

```

Carlie had a seizure 2 weeks ago. She is complaining of increased frequent headaches
Given lyme disease symptoms such as muscle ache and stiff neck will order prescription.
Meds : Topamax 50 mgs at breakfast daily. Amoxicillan 25 mg by mouth twice a day
Place MRI radiology order at RadNet

```

為了專注於診斷碼，提示中只會 DX\_NAME 使用 MEDICAL\_CONDITION 與類型相關的實體。由於不相關，會排除其他中繼資料。對於藥物實體，包含藥物名稱以及擷取的屬性。由於不相關，Amazon Comprehend Medical 的其他藥物實體中繼資料會被排除。以下是使用篩選 Amazon Comprehend Medical 結果的範例提示。提示著重於具有 DX\_NAME 類型的 MEDICAL\_CONDITION 實體。此提示旨在更精確地將診斷碼與藥物連結，並更精確地擷取醫療訂單測試：

```

<patient_note>
Carlie had a seizure 2 weeks ago. She is complaining of increased frequeunt headaches
Given lyme disease symptoms such as muscle ache and stiff neck will order
prescription.
Meds : Topamax 50 mgs at breakfast daily. Amoxicillan 25 mg by mouth twice a day

```

```
Place MRI radiology order at RadNet
</patient_note>

<detect_entity_results>
<entity>
  <text>seizure</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>headaches</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>lyme disease</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>muscle ache</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>stiff neck</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
</detect_entity_results>

<rx_results>
<entity>
  <text>Topamax</text>
  <category>MEDICATION</category>
  <type>BRAND_NAME</type>
  <attributes>
    <attribute>
      <type>FREQUENCY</type>
      <text>at breakfast daily</text>
    </attribute>
    <attribute>
      <type>DOSAGE</type>
      <text>50 mgs</text>
    </attribute>
  </attributes>
</entity>
</rx_results>
```

```

    </attribute>
    <attribute>
      <type>ROUTE_OR_MODE</type>
      <text>by mouth</text>
    </attribute>
  </attributes>
</entity>
<entity>
  <text>Amoxicillan</text>
  <category>MEDICATION</category>
  <type>GENERIC_NAME</type>
  <attributes>
    <attribute>
      <type>ROUTE_OR_MODE</type>
      <text>by mouth</text>
    </attribute>
    <attribute>
      <type>DOSAGE</type>
      <text>25 mg</text>
    </attribute>
    <attribute>
      <type>FREQUENCY</type>
      <text>twice a day</text>
    </attribute>
  </attributes>
</entity>
</rx_results>

<prompt>
Based on the patient note and the detected entities, can you please:
1. Link the diagnosis symptoms with the medications prescribed.
Provide your reasoning for the linkages.
2. Extract any entities related to medical order tests mentioned in the note.
</prompt>

```

## 使用 Amazon Comprehend Medical 套用護欄

您可以在使用產生的回應之前，使用 LLM 和 Amazon Comprehend Medical 建立護欄。您可以在未修改或經過後處理的醫療文字上執行此工作流程。使用案例包括解決受保護醫療資訊 (PHI)、偵測幻覺，或實作發佈結果的自訂政策。例如，您可以使用 Amazon Comprehend Medical 的內容來識別 PHI 資料，然後使用 LLM 來移除該 PHI 資料。

以下是來自包含 PHI 之患者記錄的資訊範例：

```
Patient name: John Doe
Patient SSN: 123-34-5678
Patient DOB: 01/01/2024
Patient address: 123 Main St, Anytown USA
Exam details: good health. Pulse is 60 bpm. needs to work on diet with BMI of 190
```

以下是提示範例，其中包含 Amazon Comprehend Medical 結果做為內容：

```
<original_text>
Patient name: John Doe
Patient SSN: 123-34-5678 Patient DOB: 01/01/2024
Patient address: 123 Main St, Anytown USA
Exam details: good health. Pulse is 60 bpm. needs to work on diet with BMI of 190
</original_text>

<comprehend_medical_phi_entities>
<entity>
  <text>John Doe</text>
  <category>PROTECTED_HEALTH_INFORMATION</category>
  <score>0.9967944025993347</score>
  <type>NAME</type>
</entity>
<entity>
  <text>123-34-5678</text>
  <category>PROTECTED_HEALTH_INFORMATION</category>
  <score>0.9998034834861755</score>
  <type>ID</type>
</entity>
<entity>
  <text>01/01/2000</text>
  <category>PROTECTED_HEALTH_INFORMATION</category>
  <score>0.9964448809623718</score>
  <type>DATE</type>
</entity>
</comprehend_medical_phi_entities>

<instructions>
Using the provided original text and the Amazon Comprehend Medical PHI entities
detected, please analyze the text to determine if it contains any additional protected
health information (PHI) beyond the entities already identified. If additional PHI is
found, please list and categorize it. If no additional PHI is found, please state that
explicitly.
In addition if PHI is found, generate updated text with the PHI removed.
```

```
</instructions>
```

## 針對醫療保健和生命科學使用案例使用大型語言模型

這說明如何將大型語言模型 (LLMs) 用於醫療保健和生命科學應用程式。有些使用案例需要使用大型語言模型來產生式 AI 功能。即使是最先進的 state-of-the-art LLMs 也有優點和限制，本節中的建議旨在協助您實現目標結果。

您可以使用決策路徑來判斷適合您使用案例的適當 LLM 解決方案，並考量領域知識和可用訓練資料等因素。此外，本節討論熱門的預先訓練醫療 LLMs 以及其選擇和使用的最佳實務。它還討論了複雜、高效能解決方案和更簡單、成本更低方法之間的權衡。

### LLM 的使用案例

Amazon Comprehend Medical 可以執行特定的 NLP 任務。如需詳細資訊，請參閱 [Amazon Comprehend Medical 的使用案例](#)。

進階醫療保健和生命科學使用案例可能需要 LLM 的邏輯和生成式 AI 功能，例如：

- 分類自訂醫療實體或文字類別
- 回答臨床問題
- 摘要醫療報告
- 從醫療資訊產生和偵測洞見

### 自訂方法

了解 LLMs 的實作方式至關重要。LLMs 通常使用數十億個參數進行訓練，包括來自許多網域的訓練資料。此訓練可讓 LLM 處理最廣義的任務。不過，當需要特定領域的知識時，通常會出現挑戰。醫療保健和生命科學領域知識的範例包括診所代碼、醫療術語，以及產生準確答案所需的健康資訊。因此，針對這些使用案例使用 LLM 為原樣（不補充網域知識的零鏡頭提示）可能會導致結果不準確。您可以使用多種常用方法來克服此挑戰：提示工程、擷取增強生成 (RAG) 和微調。

### 提示詞工程

提示詞工程是引導生成式 AI 解決方案，透過調整 LLM 的輸入來建立所需輸出的程序。透過使用相關內容建構精確的提示，可以引導模型完成需要推理的特殊醫療保健任務。有效的提示詞工程可以大幅改善醫療保健使用案例的模型效能，而不需要修改模型。如需提示詞工程的詳細資訊，請參閱 [使用 Amazon](#)

[Bedrock 實作進階提示詞工程](#) (AWS 部落格文章)。少量的提示和chain-of-thought提示是可用於提示工程的技術。

### 少量樣本提示

少量擷取提示是一種技術，您可以在要求 LLM 執行類似任務之前，提供一些所需的輸入輸出範例給 LLM。在醫療保健環境中，這種方法對於特殊任務特別有用，例如醫療實體辨識或臨床筆記摘要。透過在提示中包含 3-5 個高品質範例，您可以大幅提升模型對醫學術語和領域特定模式的理解。如需少量擷取提示的範例，請參閱 [Amazon Bedrock 中 LLMs 的少量擷取提示工程和微調](#) (AWS 部落格文章)。

例如，當您從臨床筆記擷取藥物用量時，您可以提供不同表示法樣式的範例，以協助模型辨識醫療保健專業人員記錄處方方式的變化。此方法在使用標準化文件格式或資料中存在一致模式時特別有效。

### Chain-of-thought提示

Chain-of-thought(CoT) 提示會引導 LLM 完成step-by-step推理程序。這使得它對於複雜的醫療決策支援和診斷推理任務很有價值。透過在分析臨床案例時明確指示模型「逐步思考」，您可以改善其遵循醫療推理通訊協定的能力，並減少診斷錯誤。

當臨床推理需要多個邏輯步驟，例如差異診斷或治療規劃時，這種技術會更好。不過，在處理模型訓練資料之外的高度專業化醫學知識時，或需要絕對精確度才能做出重大照護決策時，此方法會有限制。

在這些情況下，結合 CoT 與另一種方法可以產生更好的結果。其中一個選項是結合 CoT 與自我一致性提示。如需詳細資訊，請參閱[使用 Amazon Bedrock 上的自我一致性提示增強生成語言模型的效能](#) (AWS 部落格文章)。另一個選項是結合推理架構，例如 ReAct 提示和 RAG。如需詳細資訊，請參閱[使用 RAG 和 ReAct 提示開發進階生成式 AI 聊天式助理](#) (AWS 方案指引)。

### 檢索增強生成

擷取增強生成 (RAG) 是一種生成式 AI 技術，其中 LLM 會在產生回應之前參考訓練資料來源以外的授權資料來源。RAG 系統可以從知識來源擷取醫學本質資訊（例如國際疾病分類、國家藥物檔案和醫學主題標題）。這可為 LLM 提供額外的內容，以支援醫療 NLP 任務。

如 [結合 Amazon Comprehend Medical 與大型語言模型](#) 章節所述，您可以使用 RAG 方法從 Amazon Comprehend Medical 擷取內容。其他常見的知識來源包括存放在資料庫服務中的醫療網域資料，例如 Amazon OpenSearch Service、Amazon Kendra 或 Amazon Aurora。從這些知識來源擷取資訊可能會影響擷取效能，尤其是使用向量資料庫的語意查詢。

儲存和擷取特定網域知識的另一個選項是在您的 RAG 工作流程中使用 [Amazon Q Business](#)。Amazon Q Business 可以為內部文件儲存庫或面向公眾的網站編製索引（例如 [CMS.gov](#) for ICD-10 data）。然後，Amazon Q Business 可以從這些來源擷取相關資訊，再將查詢傳遞至 LLM。

有多種方式可建置自訂 RAG 工作流程。例如，從知識來源擷取資料的方法有很多。為了簡化，我們建議使用向量資料庫的常見擷取方法，例如 Amazon OpenSearch Service，將知識儲存為內嵌項目。這需要您使用內嵌模型，例如句子轉換器，來產生查詢和向量資料庫中存放知識的內嵌。

如需全受管和自訂 RAG 方法的詳細資訊，請參閱 [上的擷取增強生成選項和架構 AWS](#)。

## 微調

微調現有模型需要採用 LLM，例如 Amazon Titan、Mistral 或 Llama 模型，然後根據您的自訂資料調整模型。微調有多種技術，其中大部分只涉及修改幾個參數，而不是修改模型中的所有參數。這稱為參數效率微調 (PEFT)。如需詳細資訊，請參閱 GitHub 上的 [Hugging Face PEFT](#)。

以下是兩種常見的使用案例，您可以選擇微調醫療 NLP 任務的 LLM：

- 生成式任務 – 以解碼器為基礎的模型會執行生成式 AI 任務。AI/ML 從業人員使用 Ground Truth 資料來微調現有的 LLM。例如，您可以使用公有醫療問題回答資料集 [MedQuAD](#) 來訓練 LLM。當您調用對微調 LLM 的查詢時，您不需要 RAG 方法來提供 LLM 的其他內容。
- 內嵌 – 以編碼器為基礎的模型透過將文字轉換為數值向量來產生內嵌。這些編碼器型模型通常稱為內嵌模型。句子轉換器模型是針對句子最佳化的特定內嵌模型類型。目標是從輸入文字產生內嵌。然後，內嵌會用於語意分析或擷取任務。若要微調內嵌模型，您必須擁有可用作訓練資料的醫學知識，例如文件。這是以相似性或情緒為基礎的文字對來完成，以微調句子轉換器模型。如需詳細資訊，請參閱在 Hugging Face [上使用句子轉換器 v3 訓練和微調內嵌模型](#)。

您可以使用 [Amazon SageMaker Ground Truth](#) 來建置高品質、已標記的訓練資料集。您可以使用 Ground Truth 的已標籤資料集輸出，來訓練您自己的模型。您也可以使用輸出作為 Amazon SageMaker AI 模型的訓練資料集。如需具名實體辨識、單一標籤文字分類和多標籤文字分類的詳細資訊，請參閱 Amazon SageMaker AI 文件中的 [使用 Ground Truth 文字標籤](#)。

如需微調的詳細資訊，請參閱本指南在 [醫療保健領域微調大型語言模型](#) 中的。

## 選擇 LLM

[Amazon Bedrock](#) 是評估高效能 LLMs 的建議起點。如需詳細資訊，請參閱 [Amazon Bedrock 中支援的基礎模型](#)。您可以在 Amazon Bedrock 中使用模型評估任務，以比較來自多個輸出的輸出，然後選擇最適合您的使用案例的模型。如需詳細資訊，請參閱 [《Amazon Bedrock 文件》中的使用 Amazon Bedrock 評估選擇效能最佳的模型](#)。

有些 LLMs 對醫療網域資料的訓練有限。如果您的使用案例需要微調 Amazon Bedrock 不支援的 LLM 或 LLM，請考慮使用 [Amazon SageMaker AI](#)。在 SageMaker AI 中，您可以使用微調的 LLM，或選擇已針對醫療網域資料訓練的自訂 LLM。

下表列出已針對醫療網域資料進行訓練 LLMs。

LLM	任務	知識	Architecture
<a href="#">BioBERT</a>	資訊擷取、文字分類和具名實體辨識	來自 PubMed 的摘要、來自 PubMedCentral 的全文文章，以及一般網域知識	編碼器
<a href="#">ClinicalBERT</a>	資訊擷取、文字分類和具名實體辨識	大型、多中心資料集，以及來自電子運作狀態記錄 (EHR) 系統超過 3,000,000 個病患記錄	編碼器
<a href="#">ClinicalGPT</a>	摘要、問答和文字產生	廣泛且多樣化的醫療資料集，包括醫療記錄、領域特定知識和多回合對話諮詢	解碼器
<a href="#">GatorTron-OG</a>	摘要、問答、文字產生和資訊擷取	臨床備註和生物醫學文獻	編碼器
<a href="#">Med-BERT</a>	資訊擷取、文字分類和具名實體辨識	大型醫療文字資料集、臨床筆記、研究論文和醫療保健相關文件	編碼器
<a href="#">Med-PaLM</a>	醫學用途的問答	醫療和生物醫學文字的资料集	解碼器
<a href="#">medAlpaca</a>	問答和醫療對話任務	各種醫療文字，包括醫療閃存卡、Wiki 和對話資料集等資源	解碼器
<a href="#">BiomedBERT</a>	資訊擷取、文字分類和具名實體辨識	來自 PubMed 的獨家摘要和來自 PubMedCentral 的全文文章	編碼器

## [BioMedLM](#)

摘要、問答和文字產生 來自 PubMed 知識來源的生物醫學文獻 解碼器

以下是使用預先訓練醫療 LLMs 最佳實務：

- 了解訓練資料及其與醫療 NLP 任務的相關性。
- 識別 LLM 架構及其用途。編碼器適用於內嵌和 NLP 任務。解碼器用於產生任務。
- 評估託管預先訓練醫療 LLM 的基礎設施、效能和成本需求。
- 如果需要微調，請確保訓練資料的正確基本事實或知識。請確定您遮罩或修改任何個人身分識別資訊 (PII) 或受保護醫療資訊 (PHI)。

在知識或預期使用案例方面，實際醫療 NLP 任務可能與預先訓練 LLMs 不同。如果特定網域的 LLM 不符合您的評估基準，您可以使用自己的資料集微調 LLM，也可以訓練新的基礎模型。訓練新的基礎模型是遠大且通常昂貴的任務。對於大多數使用案例，我們建議微調現有模型。

當您使用或微調預先訓練的醫療 LLM 時，請務必解決基礎設施、安全和護欄的問題。

## 基礎設施

相較於使用 Amazon Bedrock 進行隨需或批次推論，託管預先訓練的醫療 LLMs (通常來自 Hugging Face) 需要大量資源。若要託管預先訓練的醫療 LLMs，通常會使用在具有一或多個 GPUs 的 Amazon Elastic Compute Cloud (Amazon EC2) 執行個體上執行的 Amazon SageMaker AI 映像，例如用於加速運算的 ml.g5 執行個體或 ml.inf2 執行個體 AWS Inferentia。Amazon EC2 這是因為 LLMs 會耗用大量記憶體和磁碟空間。

## 安全與護欄

根據您的業務合規要求，請考慮使用 Amazon Comprehend 和 Amazon Comprehend Medical 來遮罩或修訂訓練資料中的個人身分識別資訊 (PII) 和受保護醫療資訊 (PHI)。這有助於防止 LLM 在產生回應時使用機密資料。

我們建議您考慮並評估生成式 AI 應用程式中的偏差、公平性和幻覺。無論您是使用預先存在的 LLM 還是微調 LLM，請實作護欄以防止有害的回應。護欄是您根據生成式 AI 應用程式需求和負責任的 AI 政策自訂的保護措施。例如，您可以使用 [Amazon Bedrock 護欄](#)。

## 在醫療保健領域微調大型語言模型

本節中描述的微調方法支援遵守道德和法規準則，並促進在醫療保健中負責任地使用 AI 系統。它旨在產生準確且私密的洞見。生成式 AI 正在徹底改變醫療保健交付，但 off-the-shelf 模型在準確性至關重要且合規性不可協商的臨床環境中通常很短。使用網域特定資料微調基礎模型可彌補此差距。它可協助您建立能說醫學語言的 AI 系統，同時遵守嚴格的法規標準。不過，成功微調的途徑需要仔細導覽醫療保健的獨特挑戰：保護敏感資料、以可衡量的結果證明 AI 投資的合理性，以及在快速發展的醫療環境中維持臨床相關性。

當較輕量的方法達到限制時，微調會成為策略投資。預期準確度、延遲或營運效率的獲得將抵銷所需的大量運算和工程成本。請務必記住，基礎模型的進度速度很快，因此微調後的模型優點可能只會持續到下一個主要模型版本為止。

本節以以下兩個醫療保健客戶高影響的使用案例 AWS 為基礎討論：

- 臨床決策支援系統 – 透過了解複雜患者歷史記錄和不斷發展準則的模型來提高診斷準確性。微調可協助模型深入了解複雜的患者歷史記錄，並整合專門的指導方針 這可能會減少模型預測錯誤。不過，您需要將這些收益權衡在大型敏感資料集上進行訓練的成本，以及高風險臨床應用程式所需的基礎設施。提高的準確性和內容感知性是否會使投資合理，尤其是在頻繁發佈新模型時？
- 醫療文件分析 – 自動化處理臨床筆記、影像報告和保險文件，同時維持健康保險流通與責任法案 (HIPAA) 合規。在這裡，微調可讓模型更有效地處理唯一格式、特殊縮寫和法規要求。回報通常出現在減少的手動審核時間和改善的合規性中。不過，評估這些改進是否足夠重要，以保證微調資源是至關重要的。判斷提示工程和工作流程協調是否可以滿足您的需求。

這些真實世界案例說明從初始實驗到模型部署的微調旅程，同時在每個階段處理醫療保健的獨特需求。

### 預估成本和投資報酬率

以下是微調 LLM 時必須考量的成本因素：

- 模型大小 – 較大的模型進行微調的成本更高
- 資料集大小 – 運算成本和時間會隨著資料集大小而增加，以進行微調
- 微調策略 – 相較於完整的參數更新，參數效率方法可以降低成本

計算投資報酬率 (ROI) 時，請考慮您選擇的指標（例如準確性）的改善乘以請求量（使用模型的頻率），以及模型超過較新版本之前的預期持續時間。

此外，請考慮基本 LLM 的生命週期。新的基礎模型每 6–12 個月出現一次。如果您罕見疾病偵測器需要 8 個月的時間進行微調和驗證，在較新的模型縮小差距之前，您可能只會獲得 4 個月的卓越效能。

透過計算使用案例的成本、投資報酬率和潛在的生命週期，您可以做出資料驅動型決策。例如，如果微調您的臨床決策支援模型會導致每年數千個案例的診斷錯誤大幅減少，投資可能會快速獲得回報。相反地，如果單獨使用提示詞工程使文件分析工作流程接近您的目標準確性，則可能明智地延遲微調，直到下一代模型到達。

微調不是 one-size-fits-all。如果您決定微調，正確的方法取決於您的使用案例、資料和資源。

## 選擇微調策略

在您確定微調是醫療保健使用案例的正確方法之後，下一步是選擇最適合的微調策略。有數種方法可用。每個都有不同的醫療保健應用程式優勢和權衡。這些方法的選擇取決於您的特定目標、可用資料和資源限制。

### 訓練目標

[網域調適性預先訓練 \(DAPT\)](#) 是一種非監督式方法，涉及在大量網域特定的未標記文字（例如數百萬份醫療文件）上預先訓練模型。這種方法非常適合改善模型了解醫學專科縮寫的能力，以及放射師、神經學家和其他專業提供者所使用的術語。不過，DAPT 需要大量資料，而且不會處理特定任務輸出。

[監督式微調 \(SFT\)](#) 會使用結構化輸入輸出範例，教導模型遵循明確指示。這種方法適用於醫療文件分析工作流程，例如文件摘要或臨床編碼。指令調校是一種常見的 SFT 形式，其中模型根據範例進行訓練，其中包括搭配所需輸出的明確指示。這可增強模型了解和遵循各種使用者提示的能力。這種技術在醫療保健環境中特別有用，因為它使用特定的臨床範例來訓練模型。主要缺點是它需要仔細標記的範例。此外，微調後的模型可能會遇到沒有範例的邊緣案例。如需使用 Amazon SageMaker Jumpstart 進行微調的指示，請參閱[使用 Amazon SageMaker Jumpstart 微調 FLAN T5 XL 的說明](#) (AWS 部落格文章)。

[從人類意見回饋 \(RLHF\) 進行強化學習](#) 會根據專家意見回饋和偏好設定最佳化模型行為。使用根據人類偏好和方法訓練的獎勵模型，例如[近端政策最佳化 \(PPO\)](#) 或[直接偏好最佳化 \(DPO\)](#)，來最佳化模型，同時防止破壞性更新。RLHF 非常適合使輸出符合臨床準則，並確保建議保持在核准的通訊協定內。這種方法需要大量的臨床醫生時間來提供意見回饋，並涉及複雜的訓練管道。不過，RLHF 在醫療保健領域特別有價值，因為它有助於醫療專家塑造 AI 系統通訊和提出建議的方式。例如，臨床醫生可以提供意見回饋，以確保模型維持適當的床邊方式、知道何時表達不確定性，並保持在臨床指導方針內。PPO 等技術會根據專家意見回饋反覆最佳化模型行為，同時限制參數更新以保留核心醫學知識。這可讓模型以病患友善的語言傳達複雜的診斷，同時仍標記嚴重條件以立即進行醫療。這對於準確性和溝通風格都很重要的醫療保健至關重要。如需 RLHF 的詳細資訊，請參閱[使用人工或 AI 意見回饋的強化學習來微調大型語言模型](#) (AWS 部落格文章)。

## 實作方法

完整參數更新包括在訓練期間更新所有模型參數。此方法最適合需要深度整合患者歷史記錄、實驗室結果和不斷發展的指導方針的臨床決策支援系統。如果您的資料集不大且多樣化，缺點包括高運算成本和過度擬合的風險。

具參數效率的微調 (PEFT) 方法只會更新一部分參數，以防止過度擬合或嚴重遺失語言功能。類型包括低階適應 (LoRA)、轉接器和字首調校。PEFT 方法提供較低的運算成本、更快的訓練，非常適合用於實驗，例如根據新醫院的通訊協定或術語調整臨床決策支援模型。與完整參數更新相比，主要限制可能是效能降低。

如需微調方法的詳細資訊，請參閱 [Amazon SageMaker AI 上的進階微調方法](#) (AWS 部落格文章)。

## 建置微調資料集

微調資料集的品質和多樣性對於模型效能、安全性和預防偏差至關重要。以下是建置此資料集時需要考慮的三個關鍵領域：

- 以微調方法為基礎的磁碟區
- 網域專家的資料註釋
- 資料集的多樣性

如下表所示，微調的資料集大小需求會根據所執行微調的類型而有所不同。

微調策略	資料集大小
網域調整的預先訓練	超過 100,000 個網域文字
受監督的微調	超過 10,000 個已標記的配對
從人類意見回饋中強化學習	超過 1,000 個專家偏好設定對

您可以使用 [AWS Glue](#)、[Amazon EMR](#) 和 [Amazon SageMaker Data Wrangler](#) 自動化資料擷取和轉換程序，以策劃您擁有的資料集。如果您無法策劃夠大的資料集，您可以透過直接探索並下載資料集到您的 AWS 帳戶 [AWS Data Exchange](#)。使用任何第三方資料集之前，請先諮詢您的法律顧問。

具有領域知識的專家註釋，例如醫生、生物學家和化學家，應該是資料策劃程序的一部分，以便將醫療和生物資料的細微差別納入模型輸出。[Amazon SageMaker Ground Truth](#) 提供低程式碼使用者介面，讓專家註釋資料集。

代表人類人口的資料集對於醫療保健和生命科學微調使用案例至關重要，以防止偏差並反映真實的結果。[AWS Glue 互動式工作階段](#)或 [Amazon SageMaker 筆記本執行個體](#)提供一種強大的方法，透過使用 Jupyter 相容的筆記本反覆探索資料集和微調轉換。互動式工作階段可讓您在本地環境中使用多種熱門的整合式開發環境 (IDEs)。或者，您可以透過使用 AWS Glue 或 [Amazon SageMaker Studio](#) 筆記本 AWS 管理主控台。

## 微調模型

AWS 提供 [Amazon SageMaker AI](#) 和 [Amazon Bedrock](#) 等服務，這些服務對於成功微調至關重要。

SageMaker AI 是一種全受管的機器學習服務，可協助開發人員和資料科學家快速建置、訓練和部署 ML 模型。SageMaker AI 用於微調的三個實用功能包括：

- [SageMaker Training](#) – 全受管 ML 功能，可協助您大規模有效率地訓練各種模型
- [SageMaker JumpStart](#) – 以 SageMaker 訓練任務為基礎建置的功能，可為 ML 任務提供預先訓練的模型、內建演算法和解決方案範本
- [SageMaker HyperPod](#) – 專用基礎設施解決方案，用於基礎模型和 LLMs 的分散式訓練

Amazon Bedrock 是一項全受管服務，可透過 API 存取高效能的基礎模型，並具有內建的安全性、隱私權和可擴展性功能。此服務提供微調數個可用基礎模型的功能。如需詳細資訊，請參閱 Amazon Bedrock 文件中的 [支援模型和區域，以進行微調和持續的預先訓練](#)。

使用任一服務處理微調程序時，請考慮基本模型、微調策略和基礎設施。

## 基礎模型選擇

封閉式來源模型，例如 Anthropic Claude、Meta Llama 和 Amazon Nova，可提供具有受管合規的強大 out-of-the-box 效能，但會將微調彈性限制在供應商支援的選項，例如 Amazon Bedrock 等受管 APIs。這限制了可自訂性，尤其是受監管的醫療保健使用案例。相反地，Meta Llama 等開放原始碼模型跨 Amazon SageMaker AI 服務提供完全的控制和彈性，因此當您需要自訂、稽核或深度調整模型以滿足您的特定資料或工作流程需求時，它們是理想的選擇。

## 微調策略

Amazon Bedrock [模型自訂](#)或 Amazon SageMaker JumpStart 可以處理簡單的指令調校。複雜的 PEFT 方法，例如 LoRA 或轉接器，需要 SageMaker 訓練任務或 Amazon Bedrock 中的自訂微調功能。SageMaker HyperPod 支援非常大型模型的分散式訓練。

## 基礎設施擴展和控制

Amazon Bedrock 等全受管服務可將基礎設施管理降至最低，非常適合優先考慮易用性和合規性的組織。半受管選項，例如 SageMaker JumpStart，提供一些靈活性和更低的複雜性。這些選項適用於快速原型或使用預先建置的工作流程時。完全控制和自訂隨附 SageMaker Training 任務和 HyperPod，雖然這些任務需要更多專業知識，但最適合您需要擴展大型資料集或需要自訂管道時。

## 監控微調的模型

在醫療保健和生命科學中，監控 LLM 微調需要追蹤多個關鍵效能指標。準確度提供基準測量，但必須平衡精確度和召回率，尤其是在分類錯誤帶來重大後果的應用程式中。F1-score 有助於解決醫療資料集中常見的類別不平衡問題。如需詳細資訊，請參閱本指南中的 [評估醫療和生命科學應用程式的 LLMs](#)。

校正指標可協助您確保模型的可信度層級符合實際機率。[公平性指標](#)可協助您偵測不同患者人口統計特性的潛在偏差。

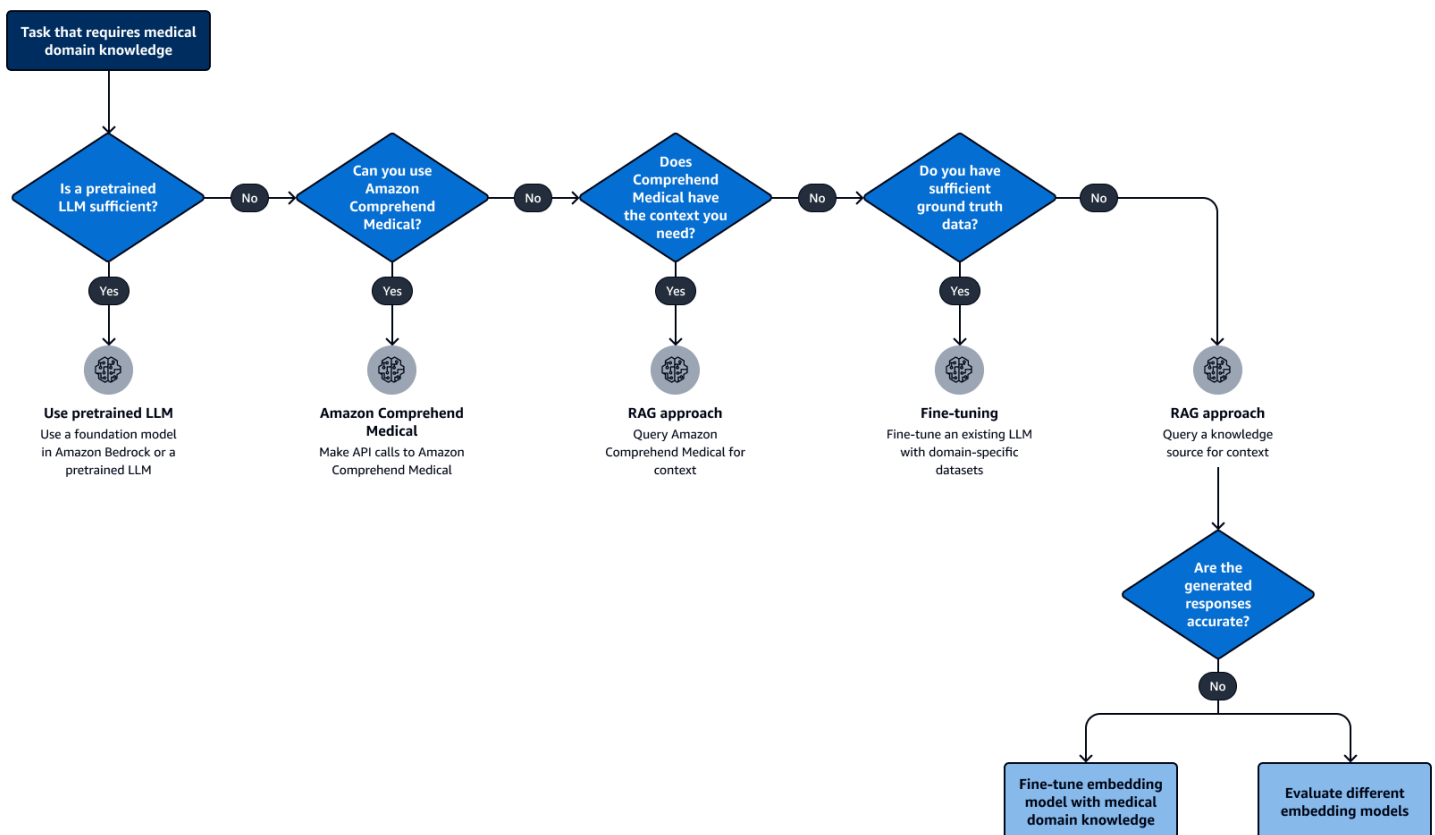
[MLflow](#) 是一種開放原始碼解決方案，可協助您追蹤微調實驗。Amazon SageMaker AI 原生支援 MLflow，可協助您以視覺化方式比較訓練執行中的指標。對於 Amazon Bedrock 上的微調任務，指標會串流到 Amazon CloudWatch，讓您可以在 CloudWatch 主控台中視覺化指標。

# 為醫療保健和生命科學選擇 NLP 方法

[醫療保健和生命科學的生成式 AI 和 NLP 方法](#) 本節說明解決醫療保健和生命科學應用程式自然語言處理 (NLP) 任務的下列方法：

- 使用 Amazon Comprehend Medical
- 在擷取增強產生 (RAG) 工作流程中結合 Amazon Comprehend Medical 與 LLM
- 使用微調的 LLM
- 使用 RAG 工作流程

透過評估醫療網域任務和使用案例的 LLMs 已知限制，您可以選擇最適合您任務的方法。下列決策樹可協助您為醫療 NLP 任務選擇 LLM 方法：



該圖顯示以下工作流程：

1. 對於醫療保健和生命科學使用案例，識別 NLP 任務是否需要特定領域知識。視需要與主題專家 (SMEs) 協調。

2. 如果您可以使用一般 LLM 或已在醫療資料集上訓練的模型，請使用 Amazon Bedrock 中可用的基礎模型或預先訓練的 LLM。如需詳細資訊，請參閱本指南中的 [選擇 LLM](#)。
3. 如果 Amazon Comprehend Medical 的實體偵測和本體連結功能處理您的使用案例，請使用 Amazon Comprehend Medical APIs。如需詳細資訊，請參閱本指南中的 [使用 Amazon Comprehend Medical](#)。
4. 有時候，Amazon Comprehend Medical 具有所需的內容，但不支援您的使用案例。例如，您可能需要不同的實體定義、接收大量結果、需要自訂實體，或需要自訂 NLP 任務。如果是這種情況，請使用 RAG 方法來查詢 Amazon Comprehend Medical 的內容。如需詳細資訊，請參閱本指南中的 [結合 Amazon Comprehend Medical 與大型語言模型](#)。
5. 如果您有足夠數量的 Ground Truth 資料，請微調現有的 LLM。如需詳細資訊，請參閱本指南中的 [自訂方法](#)。
6. 如果其他方法不符合您的 NLP 任務目標，請實作 RAG 解決方案。如需詳細資訊，請參閱本指南中的 [自訂方法](#)。
7. 實作 RAG 解決方案後，評估產生的回應是否正確。如需詳細資訊，請參閱本指南中的 [評估醫療和生命科學應用程式的 LLMs](#)。通常從 Amazon Titan Text Embeddings 模型或一般句子轉換器模型開始，例如 [all-MiniLM-L6-v2](#)。不過，由於缺乏網域內容，這些模型可能無法擷取文字的醫療術語。如有必要，請考慮下列調整：
  - a. 評估其他內嵌模型
  - b. 使用特定網域的資料集微調內嵌模型

## 業務成熟度考量事項

在為醫療保健和生命科學應用程式調整 LLM 解決方案時，業務成熟度至關重要。這些組織在實作 LLMs 時，會面臨不同的複雜程度，取決於其接受標準。通常，缺乏 AI/ML 資源的組織會投資承包商支援來建置 LLM 解決方案。在這些情況下，請務必了解下列權衡：

- 高效能以實現高成本和維護 – 您可能需要複雜的解決方案，其中包含微調或自訂 LLMs 以符合嚴格的效能標準。不過，這會產生更高的成本和維護需求。您可能需要僱用專業資源或與承包商合作，以維護這些複雜的解決方案。這可能會拖慢開發速度。
- 低成本和維護的良好效能 – 或者，您可能會發現 Amazon Bedrock 或 Amazon Comprehend Medical 等服務提供可接受的效能。雖然這些 LLMs 或方法可能會提供完美的結果，但這些解決方案通常可以提供一致、高品質的結果。這些解決方案的成本較低，並降低維護負擔。這可以加速開發。

如果更簡單、成本較低的方法持續提供符合您接受條件的高品質結果，請考慮提高的效能是否值得成本、維護和時間權衡。不過，如果較簡單的解決方案明顯缺乏目標效能，而且您的組織缺乏複雜解決方案及其維護需求的投資容量，請考慮延遲 AI/ML 開發，直到有更多資源或替代解決方案可用為止。

此外，對於依賴 LLM 的任何醫療 NLP 解決方案，我們建議您執行持續監控和評估。評估使用者隨時間的意見回饋，並實作定期評估，以確保解決方案繼續滿足您的業務目標。

# 評估醫療和生命科學應用程式的 LLMs

本節提供在醫療保健和生命科學使用案例中評估大型語言模型 (LLMs) 的需求和考量的完整概觀。

請務必使用 Ground Truth 資料和 SME 意見回饋來緩解偏差，並驗證 LLM 產生的回應的準確性。本節說明收集和策劃訓練和測試資料的最佳實務。它還可協助您實作護欄並測量資料偏差和公平性。它還討論了常見的醫療自然語言處理 (NLP) 任務，例如文字分類、具名實體辨識、文字產生及其相關評估指標。

它也提供在訓練實驗階段和後製階段執行 LLM 評估的工作流程。模型監控和 LLM 操作是此評估程序的重要元素。

## 醫療 NLP 任務的訓練和測試資料

醫療 NLP 任務通常會使用醫療體庫（例如 PubMed）或患者資訊（例如臨床患者就診備註）來分類、摘要和產生洞見。醫療專業人員、例如醫生、醫療保健管理員或技術人員，在專業知識和觀點方面各有不同。由於這些醫務人員之間的主觀性，較小的訓練和測試資料集會有偏差的風險。為了降低此風險，我們建議採用下列最佳實務：

- 使用預先訓練的 LLM 解決方案時，請確定您有足夠數量的測試資料。測試資料應與實際醫療資料非常相似。視任務而定，範圍可以從 20 到超過 100 筆記錄。
- 微調 LLM 時，請從目標醫療網域的各種 SMEs 收集足夠數量的已標記（真實）記錄。一般起點至少為 100 個高品質記錄。不過，由於任務的複雜性和您的準確性接受條件，可能需要更多記錄。
- 如果醫療使用案例需要，請實作護欄並測量資料偏差和公平性。例如，請確定 LLM 防止因病患種族描述檔而導致的誤診。如需詳細資訊，請參閱本指南中的 [安全與護欄](#) 一節。

許多 AI 研究和開發公司，例如 Anthropic，已經在其基礎模型中實作護欄，以避免毒性。您可以使用毒性偵測來檢查輸入提示和來自 LLMs 輸出回應。如需詳細資訊，請參閱 Amazon Comprehend 文件中的 [毒性偵測](#)，並參閱 Amazon Bedrock 文件中的 [護欄](#)。

在任何生成式 AI 任務中，都有幻覺的風險。您可以透過執行分類等 NLP 任務來降低此風險。您也可以使用更進階的技術，例如文字相似性指標。[BertScore](#) 是常用的文字相似性指標。如需可用於緩解幻覺之技術的詳細資訊，請參閱 [大型語言模型中幻覺緩解技術的完整調查](#)。

## 醫療 NLP 任務的指標

您可以在建立基本事實資料和 SME 提供的標籤以進行訓練和測試之後，建立可量化的指標。透過壓力測試和檢閱 LLM 結果等定性程序檢查品質，有助於快速開發。不過，指標可做為支援未來 LLM 操作的量化基準，並做為每個生產版本的效能基準。

了解醫療任務至關重要。指標通常對應至下列其中一個一般 NLP 任務：

- 文字分類 – LLM 會根據輸入提示和提供的內容，將文字分類為一或多個預先定義的類別。一個範例是使用痛苦規模來分類痛苦類別。文字分類指標的範例包括：
  - [準確性](#)
  - [精確度](#)，也稱為巨集精確度
  - [召回](#)，也稱為巨集召回
  - [F1 分數](#)，也稱為巨集 F1 分數
  - [壅塞損失](#)
- 具名實體辨識 (NER) – 也稱為文字擷取，具名實體辨識是將非結構化文字中提及的具名實體定位和分類為預先定義類別的程序。範例是從病患記錄擷取藥物的名稱。NER 指標的範例包括：
  - [準確性](#)
  - [精確度](#)
  - [召回](#)
  - [F1 分數](#)
  - [壅塞損失](#)
- 產生 – LLM 會透過處理提示和提供的內容來產生新的文字。產生包括摘要任務或問答任務。產生指標的範例包括：
  - [手勢評估的召回導向研究 \(ROUGE\)](#)
  - [使用明確 ORDERing 進行轉譯評估的指標 \(METEOR\)](#)
  - [研究下雙語評估 \(BLEU\)](#) (適用於翻譯)
  - [字串距離](#)，也稱為餘弦相似性

# 醫療保健和生命科學使用案例的常見問答集

以下是有關使用 Amazon Comprehend Medical 或 LLMs 進行醫療 NLP 任務的常見問題。

## 如何在 Amazon Comprehend Medical 和 LLM 之間進行選擇？

如果您的任務是偵測醫療文字中的醫療實體，請檢閱 [Amazon Comprehend Medical 文件](#)，以了解可以擷取哪些醫療實體，以及是否有任何[目錄](#)解決您的使用案例。如果沒有，請考慮使用 LLM。如需詳細資訊，請參閱本指南[LLM 的使用案例](#)中的 [Amazon Comprehend Medical 的使用案例](#)和。

## 如何將 Amazon Comprehend Medical 結果提供給 LLM？

您可以在 LLM 提示中將 Amazon Comprehend Medical 結果合併為內容。這可為 LLM 提供額外的醫學知識和術語。提供的內容可以改善 LLM 在實體辨識、摘要或問答等任務上的效能。本指南提供數個範例，說明如何使用 Amazon Comprehend Medical 結果建構提示。如需詳細資訊，請參閱本指南中的[結合 Amazon Comprehend Medical 與大型語言模型](#)。

## 搭配使用 Amazon Comprehend Medical 與 LLMs 時，有哪些最佳實務？

我們建議您使用 Amazon Comprehend Medical 可信度分數來篩選或排定提示中的實體的優先順序。評估其在特定資料上的效能，並驗證實體定義是否符合您的需求也很重要。結合 Amazon Comprehend Medical 與特定網域的知識來源，可進一步增強 LLM 的效能。如需詳細資訊，請參閱本指南中的[在 RAG 工作流程中使用 Amazon Comprehend Medical 的最佳實務](#)。

## 我是否應該使用預先訓練的醫療 LLM，或針對我的醫療保健使用案例微調一般 LLM？

決策取決於您的特定需求和高品質訓練資料的可用性。預先訓練的醫療 LLMs 可以提供良好的起點。不過，您可能仍然需要使用網域特定的資料進行微調。如果您有足夠的標記資料，微調一般 LLM 可能是可行的選項。如需詳細資訊，請參閱本指南[為醫療保健和生命科學選擇 NLP 方法](#)中的 [選擇 LLM](#)和。

## 如何評估醫療 NLP 任務的 LLMs 效能？

我們建議您使用量化指標，例如文字分類和具名實體辨識任務的準確性、精確度、召回率和 F1 分數。您可以針對文字產生任務使用 ROUGE 和 METEOR。擁有由主題專家標記的可靠基本事實資料，並實

作長期監控模型效能的程序非常重要。如需詳細資訊，請參閱本指南中的 [評估醫療和生命科學應用程式的 LLMs](#)。

## 高複雜性和低複雜性 LLM 解決方案之間的權衡是什麼？

微調 LLM 或建置自訂 LLM 是高度複雜的解決方案。這些方法可以改善效能，但具有更高的成本和維護需求。更簡單的解決方案，例如使用預先訓練 LLMs 或 Amazon Comprehend Medical，可能會以更低的成本和更快的開發週期提供可接受的效能。不過，這些方法可能不符合某些使用案例的嚴格準確性要求。如需詳細資訊，請參閱本指南中的 [業務成熟度考量事項](#)。

## 後續步驟和資源

本指南可協助您使用 AWS 服務 自動化生產環境中真實應用程式的醫療 NLP 和生成式 AI 任務。它說明如何使用 Amazon Comprehend Medical、Amazon Bedrock 中支援的 LLMs、預先訓練的醫療 LLMs 或微調的 LLMs，以實現您的醫療保健和生命科學業務目標。本指南說明下列方法的優點和限制：

- 獨立使用 Amazon Comprehend Medical
- 將 Amazon Comprehend Medical 結果提供給 LLM
- 在擷取增強生成 (RAG) 方法中使用預先訓練的一般 LLM 或醫療 LLM
- 微調一般 LLM 或醫療 LLM

使用本指南中的[決策樹](#)和[業務成熟度考量](#)事項，根據您組織的 AI/ML 成熟度層級，在這些方法之間進行選擇。雖然 Amazon Comprehend Medical 和 Amazon Bedrock LLMs 提供強大的功能，但只有在您正確實作和評估它們時，它們才會成功。使用本指南所述的[評估資訊和指標](#)來驗證解決方案的效能。

對於後續步驟，我們建議醫療保健 IT 經理、架構師和技術主管與 AI/ML 從業人員合作，以識別其 NLP 醫療任務。使用本指南選擇開發路徑，然後使用適當的 AWS 服務 和 功能來成功實作自動化解決方案 AWS。

## AWS 資源

- Amazon Comprehend Medical 文件：
  - [開發人員指南](#)
  - [API 參考](#)
- [Amazon Bedrock 文件](#)
  - [Amazon Bedrock 模型評估](#)
  - [在 Amazon Bedrock 中微調](#)
- [在 Amazon SageMaker AI 中微調模型](#)
- [Amazon SageMaker Ground Truth](#)
- [Amazon Comprehend 毒性偵測](#)
- [AWS 醫療保健能力合作夥伴](#)

## 其他資源

- [開啟 Medical-LLM 排行榜](#)
- [針對醫療保健的大型語言模型的調查：從資料、技術和應用程式到責任和道德](#)
- [大型語言模型是不良的醫療編碼器 — 醫學程式碼查詢基準](#)
- [從入門到專家：將醫療知識建模為一般 LLMs](#)

## 貢獻者

### 編寫

- Joe King , AWS 資深資料科學家
- Ankith Ede , AWS Solutions 架構師
- Clement Perrot , AWS 資深生成式 AI 策略師
- Jillian Forde , AWS 資深解決方案架構師
- Rajesh Sitaraman , AWS 資深交付顧問
- Ross Claytor、AWS Principal Applied Scientist
- Shivesh Ummat , AWS Solutions Architect

### 檢閱

- Dilshad Raihan Akkam Veettil , AWS 資深資料科學家
- Joseph Cottingham , AWS 深度學習架構師

### 技術寫入

- 資深 AWS 技術作者 , Lilly AbouHarb

# 文件歷史紀錄

下表描述了本指南的重大變更。如果您想收到有關未來更新的通知，可以訂閱 [RSS 摘要](#)。

變更	描述	日期
<a href="#">新區段</a>	我們在 <a href="#">醫療保健章節和提示工程章節</a> 中新增了微調大型語言模型。 <a href="https://docs.aws.amazon.com/prescriptive-guidance/latest/generative-ai-nlp-healthcare/llms.html#llm-customization-prompt-engineering">https://docs.aws.amazon.com/prescriptive-guidance/latest/generative-ai-nlp-healthcare/llms.html#llm-customization-prompt-engineering</a>	2025 年 12 月 5 日
<a href="#">初次出版</a>	—	2024 年 12 月 16 日

# AWS 規範性指引詞彙表

以下是 AWS Prescriptive Guidance 提供的策略、指南和模式中常用的術語。若要建議項目，請使用詞彙表末尾的提供意見回饋連結。

## 數字

### 7 R

將應用程式移至雲端的七種常見遷移策略。這些策略以 Gartner 在 2011 年確定的 5 R 為基礎，包括以下內容：

- 重構/重新架構 – 充分利用雲端原生功能來移動應用程式並修改其架構，以提高敏捷性、效能和可擴展性。這通常涉及移植作業系統和資料庫。範例：將您的現場部署 Oracle 資料庫 遷移至 Amazon Aurora PostgreSQL 相容版本。
- 平台轉換 (隨即重塑) – 將應用程式移至雲端，並引入一定程度的優化以利用雲端功能。範例：將內部部署 Oracle 資料庫 遷移至 中的 Amazon Relational Database Service (Amazon RDS) for Oracle AWS 雲端。
- 重新購買 (捨棄再購買) – 切換至不同的產品，通常從傳統授權移至 SaaS 模型。範例：將您的客戶關係管理 (CRM) 系統 遷移至 Salesforce.com。
- 主機轉換 (隨即轉移) – 將應用程式移至雲端，而不進行任何變更以利用雲端功能。範例：將您的現場部署 Oracle 資料庫 遷移至 中 EC2 執行個體上的 Oracle AWS 雲端。
- 重新放置 (虛擬機器監視器等級隨即轉移) – 將基礎設施移至雲端，無需購買新硬體、重寫應用程式或修改現有操作。您可以將伺服器從內部部署平台遷移到相同平台的雲端服務。範例：將 Microsoft Hyper-V 應用程式 遷移至 AWS。
- 保留 (重新檢視) – 將應用程式保留在來源環境中。其中可能包括需要重要重構的應用程式，且您希望將該工作延遲到以後，以及您想要保留的舊版應用程式，因為沒有業務理由來進行遷移。
- 淘汰 – 解除委任或移除來源環境中不再需要的應用程式。

## A

### ABAC

請參閱 [屬性型存取控制](#)。

## 抽象服務

請參閱 [受管服務](#)。

## ACID

請參閱 [原子性、一致性、隔離性、持久性](#)。

## 主動-主動式遷移

一種資料庫遷移方法，其中來源和目標資料庫保持同步 (透過使用雙向複寫工具或雙重寫入操作)，且兩個資料庫都在遷移期間處理來自連接應用程式的交易。此方法支援小型、受控制批次的遷移，而不需要一次性切換。它更靈活，但比 [主動-被動遷移](#) 需要更多的工作。

## 主動-被動式遷移

一種資料庫遷移方法，其中來源和目標資料庫保持同步，但只有來源資料庫會在資料複寫至目標資料庫時處理來自連線應用程式的交易。目標資料庫在遷移期間不接受任何交易。

## 彙總函數

在一組資料列上運作的 SQL 函數，會計算群組的單一傳回值。彙總函數的範例包括 SUM 和 MAX。

## AI

請參閱 [人工智慧](#)。

## AIOps

請參閱 [人工智慧操作](#)。

## 匿名化

永久刪除資料集中個人資訊的程序。匿名化有助於保護個人隱私權。匿名資料不再被視為個人資料。

## 反模式

經常用於經常性問題的解決方案，其中解決方案具有反生產力、無效或比替代解決方案更有效。

## 應用程式控制

一種安全方法，僅允許使用核准的應用程式，以協助保護系統免受惡意軟體攻擊。

## 應用程式組合

有關組織使用的每個應用程式的詳細資訊的集合，包括建置和維護應用程式的成本及其商業價值。此資訊是 [產品組合探索和分析程序](#) 的關鍵，有助於識別要遷移、現代化和優化的應用程式並排定其優先順序。

## 人工智慧 (AI)

電腦科學領域，致力於使用運算技術來執行通常與人類相關的認知功能，例如學習、解決問題和識別模式。如需詳細資訊，請參閱[什麼是人工智慧？](#)

## 人工智慧操作 (AIOps)

使用機器學習技術解決操作問題、減少操作事件和人工干預以及提高服務品質的程序。如需有關如何在 AWS 遷移策略中使用 AIOps 的詳細資訊，請參閱[操作整合指南](#)。

## 非對稱加密

一種加密演算法，它使用一對金鑰：一個用於加密的公有金鑰和一個用於解密的私有金鑰。您可以共用公有金鑰，因為它不用於解密，但對私有金鑰存取應受到高度限制。

## 原子性、一致性、隔離性、持久性 (ACID)

一組軟體屬性，即使在出現錯誤、電源故障或其他問題的情況下，也能確保資料庫的資料有效性和操作可靠性。

## 屬性型存取控制 (ABAC)

根據使用者屬性 (例如部門、工作職責和團隊名稱) 建立精細許可的實務。如需詳細資訊，請參閱《AWS Identity and Access Management (IAM) 文件》中的[ABAC for AWS](#)。

## 授權資料來源

存放主要版本資料的位置，被視為最可靠的資訊來源。您可以將授權資料來源中的資料複製到其他位置，以處理或修改資料，例如匿名、修訂或假名化資料。

## 可用區域

中的不同位置 AWS 區域，可隔離其他可用區域中的故障，並提供相同區域中其他可用區域的低成本、低延遲網路連線能力。

## AWS 雲端採用架構 (AWS CAF)

的指導方針和最佳實務架構 AWS，可協助組織制定高效且有效的計劃，以成功地移至雲端。AWS CAF 將指導方針組織到六個重點領域：業務、人員、治理、平台、安全和營運。業務、人員和控管層面著重於業務技能和程序；平台、安全和操作層面著重於技術技能和程序。例如，人員層面針對處理人力資源 (HR)、人員配備功能和人員管理的利害關係人。因此，AWS CAF 為人員開發、訓練和通訊提供指引，協助組織做好成功採用雲端的準備。如需詳細資訊，請參閱[AWS CAF 網站](#)和[AWS CAF 白皮書](#)。

## AWS 工作負載資格架構 (AWS WQF)

一種工具，可評估資料庫遷移工作負載、建議遷移策略，並提供工作預估值。AWS WQF 隨附於 AWS Schema Conversion Tool (AWS SCT)。它會分析資料庫結構描述和程式碼物件、應用程式程式碼、相依性和效能特性，並提供評估報告。

## B

### 錯誤的機器人

旨在中斷或傷害個人或組織的[機器人](#)。

### BCP

請參閱[業務持續性規劃](#)。

### 行為圖

資源行為的統一互動式檢視，以及一段時間後的互動。您可以將行為圖與 Amazon Detective 搭配使用來檢查失敗的登入嘗試、可疑的 API 呼叫和類似動作。如需詳細資訊，請參閱偵測文件中的[行為圖中的資料](#)。

### 大端序系統

首先儲存最高有效位元組的系統。另請參閱 [Endianness](#)。

### 二進制分類

預測二進制結果的過程 (兩個可能的類別之一)。例如，ML 模型可能需要預測諸如「此電子郵件是否是垃圾郵件？」等問題或「產品是書還是汽車？」

### Bloom 篩選條件

一種機率性、記憶體高效的資料結構，用於測試元素是否為集的成員。

### 藍/綠部署

一種部署策略，您可以在其中建立兩個不同但相同的環境。您可以在一個環境（藍色）中執行目前的應用程式版本，並在另一個環境（綠色）中執行新的應用程式版本。此策略可協助您快速復原，並將影響降至最低。

### 機器人

透過網際網路執行自動化任務並模擬人類活動或互動的軟體應用程式。有些機器人有用或有益，例如在網際網路上編製資訊索引的 Web 爬蟲程式。有些其他機器人稱為惡意機器人，旨在中斷或傷害個人或組織。

## 殭屍網路

受到[惡意軟體](#)感染且受單一方控制之[機器人](#)的網路，稱為機器人繼承器或機器人運算子。殭屍網路是擴展機器人及其影響的最佳已知機制。

## 分支

程式碼儲存庫包含的區域。儲存庫中建立的第一個分支是主要分支。您可以從現有分支建立新分支，然後在新分支中開發功能或修正錯誤。您建立用來建立功能的分支通常稱為功能分支。當準備好發佈功能時，可以將功能分支合併回主要分支。如需詳細資訊，請參閱[關於分支](#) (GitHub 文件)。

## 碎片存取

在特殊情況下，並透過核准的程序，讓使用者快速取得他們通常無權存取 AWS 帳戶 之 的存取權。如需詳細資訊，請參閱 Well-Architected 指南中的 AWS [實作打破玻璃程序](#) 指標。

## 棕地策略

環境中的現有基礎設施。對系統架構採用棕地策略時，可以根據目前系統和基礎設施的限制來設計架構。如果正在擴展現有基礎設施，則可能會混合棕地和[綠地](#)策略。

## 緩衝快取

儲存最常存取資料的記憶體區域。

## 業務能力

業務如何創造價值 (例如，銷售、客戶服務或營銷)。業務能力可驅動微服務架構和開發決策。如需詳細資訊，請參閱在 [AWS 上執行容器化微服務](#) 白皮書的 [圍繞業務能力進行組織](#) 部分。

## 業務連續性規劃 (BCP)

一種解決破壞性事件 (如大規模遷移) 對營運的潛在影響並使業務能夠快速恢復營運的計畫。

# C

## CAF

請參閱[AWS 雲端採用架構](#)。

## Canary 部署

版本對最終使用者的緩慢和增量版本。當您有信心時，您可以部署新版本並完全取代目前的版本。

## CCoE

請參閱 [Cloud Center of Excellence](#)。

## CDC

請參閱[變更資料擷取](#)。

### 變更資料擷取 (CDC)

追蹤對資料來源 (例如資料庫表格) 的變更並記錄有關變更的中繼資料的程序。您可以將 CDC 用於各種用途，例如稽核或複寫目標系統中的變更以保持同步。

### 混沌工程

故意引入故障或破壞性事件，以測試系統的彈性。您可以使用 [AWS Fault Injection Service \(AWS FIS\)](#) 執行試驗，為您的 AWS 工作負載帶來壓力，並評估其回應。

## CI/CD

請參閱[持續整合和持續交付](#)。

### 分類

有助於產生預測的分類程序。用於分類問題的 ML 模型可預測離散值。離散值永遠彼此不同。例如，模型可能需要評估影像中是否有汽車。

### 用戶端加密

在目標 AWS 服務接收資料之前，在本機加密資料。

### 雲端卓越中心 (CCoE)

一個多學科團隊，可推動整個組織的雲端採用工作，包括開發雲端最佳實務、調動資源、制定遷移時間表以及領導組織進行大規模轉型。如需詳細資訊，請參閱 AWS 雲端企業策略部落格上的 [CCoE 文章](#)。

### 雲端運算

通常用於遠端資料儲存和 IoT 裝置管理的雲端技術。雲端運算通常連接到[邊緣運算](#)技術。

### 雲端操作模型

在 IT 組織中，用於建置、成熟和最佳化一或多個雲端環境的操作模型。如需詳細資訊，請參閱[建置您的雲端操作模型](#)。

### 採用雲端階段

組織在遷移至時通常會經歷的四個階段 AWS 雲端：

- 專案 – 執行一些與雲端相關的專案以進行概念驗證和學習用途
- 基礎 – 進行基礎投資以擴展雲端採用 (例如，建立登陸區域、定義 CCoE、建立營運模型)

- 遷移 – 遷移個別應用程式
- 重塑 – 優化產品和服務，並在雲端中創新

部落格文章中的 Stephen Orban 定義了這些階段：AWS 雲端 企業策略部落格上的 [邁向雲端優先之旅和採用階段](#)。如需有關它們如何與 AWS 遷移策略相關的詳細資訊，請參閱 [遷移整備指南](#)。

## CMDB

請參閱 [組態管理資料庫](#)。

### 程式碼儲存庫

透過版本控制程序來儲存及更新原始程式碼和其他資產 (例如文件、範例和指令碼) 的位置。常見的雲端儲存庫包括 GitHub 或 Bitbucket Cloud。程式碼的每個版本都稱為分支。在微服務結構中，每個儲存庫都專用於單個功能。單一 CI/CD 管道可以使用多個儲存庫。

### 冷快取

一種緩衝快取，它是空的、未填充的，或者包含過時或不相關的資料。這會影響效能，因為資料庫執行個體必須從主記憶體或磁碟讀取，這比從緩衝快取讀取更慢。

### 冷資料

很少存取且通常是歷史資料的資料。查詢這類資料時，通常可接受慢查詢。將此資料移至效能較低且成本較低的儲存層或類別，可以降低成本。

### 電腦視覺 (CV)

使用機器學習從數位影像和影片等視覺化格式分析和擷取資訊的 [AI](#) 欄位。例如，Amazon SageMaker AI 提供 CV 的影像處理演算法。

### 組態偏離

對於工作負載，組態會從預期狀態變更。這可能會導致工作負載變得不合規，而且通常是漸進和無意的。

### 組態管理資料庫 (CMDB)

儲存和管理有關資料庫及其 IT 環境的資訊的儲存庫，同時包括硬體和軟體元件及其組態。您通常在遷移的產品組合探索和分析階段使用 CMDB 中的資料。

### 一致性套件

您可以組合的 AWS Config 規則和修補動作集合，以自訂您的合規和安全檢查。您可以使用 YAML 範本，將一致性套件部署為 AWS 帳戶 和 區域中或整個組織的單一實體。如需詳細資訊，請參閱 AWS Config 文件中的 [一致性套件](#)。

## 持續整合和持續交付 (CI/CD)

自動化軟體發行程度的來源、建置、測試、暫存和生產階段的程序。CI/CD 通常被描述為管道。CI/CD 可協助您將程序自動化、提升生產力、改善程式碼品質以及加快交付速度。如需詳細資訊，請參閱[持續交付的優點](#)。CD 也可表示持續部署。如需詳細資訊，請參閱[持續交付與持續部署](#)。

## CV

請參閱[電腦視覺](#)。

## D

### 靜態資料

網路中靜止的資料，例如儲存中的資料。

### 資料分類

根據重要性和敏感性來識別和分類網路資料的程序。它是所有網路安全風險管理策略的關鍵組成部分，因為它可以協助您確定適當的資料保護和保留控制。資料分類是 AWS Well-Architected Framework 中安全支柱的元件。如需詳細資訊，請參閱[資料分類](#)。

### 資料偏離

生產資料與用於訓練 ML 模型的資料之間有意義的變化，或輸入資料隨時間有意義的變更。資料偏離可以降低 ML 模型預測的整體品質、準確性和公平性。

### 傳輸中的資料

在您的網路中主動移動的資料，例如在網路資源之間移動。

### 資料網格

架構架構，提供分散式、分散式資料擁有權與集中式管理。

### 資料最小化

僅收集和處理嚴格必要資料的原則。在 中實作資料最小化 AWS 雲端 可以降低隱私權風險、成本和分析碳足跡。

### 資料周邊

AWS 環境中的一組預防性防護機制，可協助確保只有信任的身分才能從預期的網路存取信任的資源。如需詳細資訊，請參閱[在上建置資料周邊 AWS](#)。

## 資料預先處理

將原始資料轉換成 ML 模型可輕鬆剖析的格式。預處理資料可能意味著移除某些欄或列，並解決遺失、不一致或重複的值。

## 資料來源

在整個生命週期中追蹤資料的原始伺服器 and 歷史記錄的程序，例如資料的產生、傳輸和儲存方式。

## 資料主體

正在收集和處理其資料的個人。

## 資料倉儲

支援商業智慧的資料管理系統，例如分析。資料倉儲通常包含大量歷史資料，通常用於查詢和分析。

## 資料庫定義語言 (DDL)

用於建立或修改資料庫中資料表和物件之結構的陳述式或命令。

## 資料庫處理語言 (DML)

用於修改 (插入、更新和刪除) 資料庫中資訊的陳述式或命令。

## DDL

請參閱[資料庫定義語言](#)。

## 深度整體

結合多個深度學習模型進行預測。可以使用深度整體來獲得更準確的預測或估計預測中的不確定性。

## 深度學習

一個機器學習子領域，它使用多層人工神經網路來識別感興趣的輸入資料與目標變數之間的對應關係。

## 深度防禦

這是一種資訊安全方法，其中一系列的安全機制和控制項會在整個電腦網路中精心分層，以保護網路和其中資料的機密性、完整性和可用性。當您在上採用此策略時 AWS，您可以在 AWS Organizations 結構的不同層新增多個控制項，以協助保護資源。例如，defense-in-depth 方法可能會結合多重要素驗證、網路分割和加密。

## 委派的管理員

在中 AWS Organizations，相容的服務可以註冊 AWS 成員帳戶，以管理組織的帳戶和管理該服務的許可。此帳戶稱為該服務的委派管理員。如需詳細資訊和相容服務清單，請參閱 AWS Organizations 文件中的 [可搭配 AWS Organizations 運作的服務](#)。

## deployment

在目標環境中提供應用程式、新功能或程式碼修正的程序。部署涉及在程式碼庫中實作變更，然後在應用程式環境中建置和執行該程式碼庫。

## 開發環境

請參閱 [環境](#)。

## 偵測性控制

一種安全控制，用於在事件發生後偵測、記錄和提醒。這些控制是第二道防線，提醒您注意繞過現有預防性控制的安全事件。如需詳細資訊，請參閱在 AWS 上實作安全控制中的 [偵測性控制](#)。

## 開發值串流映射 (DVSM)

一種程序，用於識別對軟體開發生命週期中的速度和品質造成負面影響的限制並排定優先順序。DVSM 擴展了最初專為精簡製造實務設計的價值串流映射程序。它著重於透過軟體開發程序建立和移動價值所需的步驟和團隊。

## 數位分身

真實世界系統的虛擬呈現，例如建築物、工廠、工業設備或生產線。數位分身支援預測性維護、遠端監控和生產最佳化。

## 維度資料表

在 [星星結構描述](#) 中，較小的資料表包含有關事實資料表中量化資料的資料屬性。維度資料表屬性通常是文字欄位或離散數字，其行為類似於文字。這些屬性通常用於查詢限制、篩選和結果集標記。

## 災難

防止工作負載或系統在其主要部署位置實現其業務目標的事件。這些事件可能是自然災難、技術故障或人為動作的結果，例如意外設定錯誤或惡意軟體攻擊。

## 災難復原 (DR)

您用來將 [災難](#) 造成的停機時間和資料遺失降至最低的策略和程序。如需詳細資訊，請參閱 AWS Well-Architected Framework 中的 [上工作負載的災難復原 AWS：雲端中的復原](#)。

## DML

請參閱[資料庫處理語言](#)。

### 領域驅動的設計

一種開發複雜軟體系統的方法，它會將其元件與每個元件所服務的不斷發展的領域或核心業務目標相關聯。Eric Evans 在其著作 *Domain-Driven Design: Tackling Complexity in the Heart of Software* (Boston: Addison-Wesley Professional, 2003) 中介紹了這一概念。如需有關如何將領域驅動的設計與 strangler fig 模式搭配使用的資訊，請參閱[使用容器和 Amazon API Gateway 逐步現代化舊版 Microsoft ASP.NET \(ASMX\) Web 服務](#)。

## DR

請參閱[災難復原](#)。

### 偏離偵測

追蹤與基準組態的偏差。例如，您可以使用 AWS CloudFormation 來偵測系統資源中的偏離，也可以使用 AWS Control Tower 來[偵測登陸區域中可能影響控管要求合規性的變更](#)。<https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/using-cfn-stack-drift.html>

## DVSM

請參閱[開發值串流映射](#)。

## E

### EDA

請參閱[探索性資料分析](#)。

### EDI

請參閱[電子資料交換](#)。

### 邊緣運算

提升 IoT 網路邊緣智慧型裝置運算能力的技術。與[雲端運算](#)相比，邊緣運算可以減少通訊延遲並改善回應時間。

### 電子資料交換 (EDI)

在組織之間自動交換商業文件。如需詳細資訊，請參閱[什麼是電子資料交換](#)。

## 加密

將人類可讀取的純文字資料轉換為加密文字的運算程序。

### 加密金鑰

由加密演算法產生的隨機位元的加密字串。金鑰長度可能有所不同，每個金鑰的設計都是不可預測且唯一的。

### 端序

位元組在電腦記憶體中的儲存順序。大端序系統首先儲存最高有效位元組。小端序系統首先儲存最低有效位元組。

### 端點

請參閱 [服務端點](#)。

### 端點服務

您可以在虛擬私有雲端 (VPC) 中託管以與其他使用者共用的服務。您可以使用 [建立端點服務](#)，AWS PrivateLink 並將許可授予其他 AWS 帳戶 或 AWS Identity and Access Management (IAM) 委託人。這些帳戶或主體可以透過建立介面 VPC 端點私下連接至您的端點服務。如需詳細資訊，請參閱 Amazon Virtual Private Cloud (Amazon VPC) 文件中的 [建立端點服務](#)。

### 企業資源規劃 (ERP)

一種系統，可自動化和管理企業的關鍵業務流程（例如會計、[MES](#) 和專案管理）。

### 信封加密

使用另一個加密金鑰對某個加密金鑰進行加密的程序。如需詳細資訊，請參閱 [\(\) 文件中的信封加密](#)。AWS Key Management Service AWS KMS

### 環境

執行中應用程式的執行個體。以下是雲端運算中常見的環境類型：

- 開發環境 – 執行中應用程式的執行個體，只有負責維護應用程式的核心團隊才能使用。開發環境用來測試變更，然後再將開發環境提升到較高的環境。此類型的環境有時稱為測試環境。
- 較低的環境 – 應用程式的所有開發環境，例如用於初始建置和測試的開發環境。
- 生產環境 – 最終使用者可以存取的執行中應用程式的執行個體。在 CI/CD 管道中，生產環境是最後一個部署環境。
- 較高的環境 – 核心開發團隊以外的使用者可存取的所有環境。這可能包括生產環境、生產前環境以及用於使用者接受度測試的環境。

## epic

在敏捷方法中，有助於組織工作並排定工作優先順序的功能類別。epic 提供要求和實作任務的高層級描述。例如，AWS CAF 安全概念包括身分和存取管理、偵測控制、基礎設施安全、資料保護和事件回應。如需有關 AWS 遷移策略中的 Epic 的詳細資訊，請參閱[計畫實作指南](#)。

## ERP

請參閱[企業資源規劃](#)。

## 探索性資料分析 (EDA)

分析資料集以了解其主要特性的過程。您收集或彙總資料，然後執行初步調查以尋找模式、偵測異常並檢查假設。透過計算摘要統計並建立資料可視化來執行 EDA。

## F

### 事實資料表

[星狀結構描述](#)中的中央資料表。它存放有關業務操作的量化資料。一般而言，事實資料表包含兩種類型的資料欄：包含度量的資料，以及包含維度資料表外部索引鍵的資料欄。

### 快速失敗

一種使用頻繁和增量測試來縮短開發生命週期的理念。這是敏捷方法的關鍵部分。

### 故障隔離界限

在中 AWS 雲端，像是可用區域 AWS 區域、控制平面或資料平面等邊界會限制故障的影響，並有助於改善工作負載的彈性。如需詳細資訊，請參閱[AWS 故障隔離界限](#)。

### 功能分支

請參閱[分支](#)。

### 特徵

用來進行預測的輸入資料。例如，在製造環境中，特徵可能是定期從製造生產線擷取的影像。

### 功能重要性

特徵對於模型的預測有多重要。這通常表示為可以透過各種技術來計算的數值得分，例如 Shapley Additive Explanations (SHAP) 和積分梯度。如需詳細資訊，請參閱[機器學習模型可解譯性 AWS](#)。

## 特徵轉換

優化 ML 程序的資料，包括使用其他來源豐富資料、調整值、或從單一資料欄位擷取多組資訊。這可讓 ML 模型從資料中受益。例如，如果將「2021-05-27 00:15:37」日期劃分為「2021」、「五月」、「週四」和「15」，則可以協助學習演算法學習與不同資料元件相關聯的細微模式。

### 少量擷取提示

在要求 [LLM](#) 執行類似的任務之前，提供少量示範任務和所需輸出的範例給 LLM。此技術是內容內學習的應用程式，其中模型會從內嵌在提示中的範例 (快照) 中學習。對於需要特定格式、推理或網域知識的任務，少量的提示非常有效。另請參閱[零鏡頭提示](#)。

## FGAC

請參閱[精細存取控制](#)。

### 精細存取控制 (FGAC)

使用多個條件來允許或拒絕存取請求。

### 閃切遷移

一種資料庫遷移方法，透過[變更資料擷取](#)使用連續資料複寫，以盡可能在最短的時間內遷移資料，而不是使用分階段方法。目標是將停機時間降至最低。

## FM

請參閱[基礎模型](#)。

### 基礎模型 (FM)

大型深度學習神經網路，已在廣義和未標記資料的大量資料集上進行訓練。FMs 能夠執行各種一般任務，例如了解語言、產生文字和影像，以及以自然語言交談。如需詳細資訊，請參閱[什麼是基礎模型](#)。

## G

### 生成式 AI

已針對大量資料進行訓練的 [AI](#) 模型子集，可使用簡單的文字提示建立新的內容和成品，例如影像、影片、文字和音訊。如需詳細資訊，請參閱[什麼是生成式 AI](#)。

### 地理封鎖

請參閱[地理限制](#)。

## 地理限制 (地理封鎖)

Amazon CloudFront 中的選項，可防止特定國家/地區的使用者存取內容分發。您可以使用允許清單或封鎖清單來指定核准和禁止的國家/地區。如需詳細資訊，請參閱 CloudFront 文件中的[限制內容的地理分佈](#)。

## Gitflow 工作流程

這是一種方法，其中較低和較高環境在原始碼儲存庫中使用不同分支。Gitflow 工作流程被視為舊版，而以[幹線為基礎的工作流程](#)是現代、偏好的方法。

## 黃金影像

系統或軟體的快照，做為部署該系統或軟體新執行個體的範本。例如，在製造中，黃金映像可用於在多個裝置上佈建軟體，並有助於提高裝置製造操作的速度、可擴展性和生產力。

## 綠地策略

新環境中缺乏現有基礎設施。對系統架構採用綠地策略時，可以選擇所有新技術，而不會限制與現有基礎設施的相容性，也稱為[棕地](#)。如果正在擴展現有基礎設施，則可能會混合棕地和綠地策略。

## 防護機制

有助於跨組織單位 (OU) 來管控資源、政策和合規的高層級規則。預防性防護機制會強制執行政策，以確保符合合規標準。透過使用服務控制政策和 IAM 許可界限來將其實施。偵測性防護機制可偵測政策違規和合規問題，並產生提醒以便修正。它們是透過使用 AWS Config、AWS Security Hub、CSPM、Amazon GuardDuty、Amazon Inspector、AWS Trusted Advisor 和自訂 AWS Lambda 檢查來實施。

# H

## HA

請參閱[高可用性](#)。

## 異質資料庫遷移

將來源資料庫遷移至使用不同資料庫引擎的目標資料庫 (例如，Oracle 至 Amazon Aurora)。異質遷移通常是重新架構工作的一部分，而轉換結構描述可能是一項複雜任務。[AWS 提供有助於結構描述轉換的 AWS SCT](#)。

## 高可用性 (HA)

在遇到挑戰或災難時，工作負載能夠在不介入的情況下持續運作。HA 系統的設計目的是自動容錯移轉、持續提供高品質的效能，以及處理不同的負載和故障，並將效能影響降至最低。

## 歷史現代化

一種方法，用於現代化和升級操作技術 (OT) 系統，以更好地滿足製造業的需求。歷史資料是一種資料庫，用於從工廠中的各種來源收集和存放資料。

### 保留資料

從用於訓練機器學習模型的資料集中保留的部分歷史標記資料。您可以使用保留資料，透過比較模型預測與保留資料來評估模型效能。

### 異質資料庫遷移

將您的來源資料庫遷移至共用相同資料庫引擎的目標資料庫 (例如，Microsoft SQL Server 至 Amazon RDS for SQL Server)。同質遷移通常是主機轉換或平台轉換工作的一部分。您可以使用原生資料庫公用程式來遷移結構描述。

### 熱資料

經常存取的資料，例如即時資料或最近的轉譯資料。此資料通常需要高效能儲存層或類別，才能提供快速的查詢回應。

### 修補程序

緊急修正生產環境中的關鍵問題。由於其緊迫性，通常會在典型 DevOps 發行工作流程之外執行修補程式。

### 超級護理期間

在切換後，遷移團隊在雲端管理和監控遷移的應用程式以解決任何問題的時段。通常，此期間的長度為 1-4 天。在超級護理期間結束時，遷移團隊通常會將應用程式的責任轉移給雲端營運團隊。

## I

### IaC

將[基礎設施視為程式碼](#)。

### 身分型政策

連接至一或多個 IAM 主體的政策，可定義其在 AWS 雲端環境中的許可。

### 閒置應用程式

90 天期間 CPU 和記憶體平均使用率在 5% 至 20% 之間的應用程式。在遷移專案中，通常會淘汰這些應用程式或將其保留在內部部署。

## IIoT

請參閱[工業物聯網](#)。

### 不可變的基礎設施

為生產工作負載部署新基礎設施的模型，而不是更新、修補或修改現有的基礎設施。不可變基礎設施本質上比[可變基礎設施](#)更一致、可靠且可預測。如需詳細資訊，請參閱 AWS Well-Architected Framework [中的使用不可變基礎設施的部署](#)最佳實務。

### 傳入 (輸入) VPC

在 AWS 多帳戶架構中，接受、檢查和路由來自應用程式外部之網路連線的 VPC。[AWS 安全參考架構](#)建議您使用傳入、傳出和檢查 VPC 來設定網路帳戶，以保護應用程式與更廣泛的網際網路之間的雙向介面。

### 增量遷移

一種切換策略，您可以在其中將應用程式分成小部分遷移，而不是執行單一、完整的切換。例如，您最初可能只將一些微服務或使用者移至新系統。確認所有項目都正常運作之後，您可以逐步移動其他微服務或使用者，直到可以解除委任舊式系統。此策略可降低與大型遷移關聯的風險。

## 工業 4.0

由 [Klaus Schwab](#) 於 2016 年推出的術語，透過連線能力、即時資料、自動化、分析和 AI/ML 的進展，指製造程序的現代化。

### 基礎設施

應用程式環境中包含的所有資源和資產。

### 基礎設施即程式碼 (IaC)

透過一組組態檔案來佈建和管理應用程式基礎設施的程序。IaC 旨在協助您集中管理基礎設施，標準化資源並快速擴展，以便新環境可重複、可靠且一致。

### 工業物聯網 (IIoT)

在製造業、能源、汽車、醫療保健、生命科學和農業等產業領域使用網際網路連線的感測器和裝置。如需詳細資訊，請參閱[建立工業物聯網 \(IIoT\) 數位轉型策略](#)。

### 檢查 VPC

在 AWS 多帳戶架構中，集中式 VPC 可管理 VPCs ( 在相同或不同的 中 AWS 區域)、網際網路和內部部署網路之間的網路流量檢查。[AWS 安全參考架構](#)建議您使用傳入、傳出和檢查 VPC 來設定網路帳戶，以保護應用程式與更廣泛的網際網路之間的雙向介面。

## 物聯網 (IoT)

具有內嵌式感測器或處理器的相連實體物體網路，其透過網際網路或本地通訊網路與其他裝置和系統進行通訊。如需詳細資訊，請參閱[什麼是 IoT？](#)

### 可解釋性

機器學習模型的一個特徵，描述了人類能夠理解模型的預測如何依賴於其輸入的程度。如需詳細資訊，請參閱[的機器學習模型可解釋性 AWS](#)。

## IoT

請參閱[物聯網](#)。

## IT 資訊庫 (ITIL)

一組用於交付 IT 服務並使這些服務與業務需求保持一致的最佳實務。ITIL 為 ITSM 提供了基礎。

## IT 服務管理 (ITSM)

與組織的設計、實作、管理和支援 IT 服務關聯的活動。如需有關將雲端操作與 ITSM 工具整合的資訊，請參閱[操作整合指南](#)。

## ITIL

請參閱[IT 資訊庫](#)。

## ITSM

請參閱[IT 服務管理](#)。

## L

## 標籤型存取控制 (LBAC)

強制存取控制 (MAC) 的實作，其中使用者和資料本身都會獲得明確指派的安全標籤值。使用者安全標籤和資料安全標籤之間的交集會決定使用者可以看到哪些資料列和資料欄。

## 登陸區域

登陸區域是架構良好的多帳戶 AWS 環境，可擴展且安全。這是一個起點，您的組織可以從此起點快速啟動和部署工作負載與應用程式，並對其安全和基礎設施環境充滿信心。如需有關登陸區域的詳細資訊，請參閱[設定安全且可擴展的多帳戶 AWS 環境](#)。

## 大型語言模型 (LLM)

預先訓練大量資料的深度學習 [AI](#) 模型。LLM 可以執行多個任務，例如回答問題、摘要文件、將文字翻譯成其他語言，以及完成句子。如需詳細資訊，請參閱[什麼是 LLMs](#)。

### 大型遷移

遷移 300 部或更多伺服器。

### LBAC

請參閱[標籤型存取控制](#)。

### 最低權限

授予執行任務所需之最低許可的安全最佳實務。如需詳細資訊，請參閱 IAM 文件中的[套用最低權限許可](#)。

### 隨即轉移

請參閱 [7 個 R](#)。

### 小端序系統

首先儲存最低有效位元組的系統。另請參閱 [Endianness](#)。

## LLM

請參閱[大型語言模型](#)。

### 較低的環境

請參閱 [環境](#)。

## M

### 機器學習 (ML)

一種使用演算法和技術進行模式識別和學習的人工智慧。機器學習會進行分析並從記錄的資料 (例如物聯網 (IoT) 資料) 中學習，以根據模式產生統計模型。如需詳細資訊，請參閱[機器學習](#)。

### 主要分支

請參閱[分支](#)。

## 惡意軟體

旨在危及電腦安全或隱私權的軟體。惡意軟體可能會中斷電腦系統、洩露敏感資訊，或取得未經授權的存取。惡意軟體的範例包括病毒、蠕蟲、勒索軟體、特洛伊木馬、間諜軟體和鍵盤記錄器。

## 受管服務

AWS 服務會 AWS 操作基礎設施層、作業系統和平台，而您會存取端點來存放和擷取資料。Amazon Simple Storage Service (Amazon S3) 和 Amazon DynamoDB 是受管服務的範例。這些也稱為抽象服務。

## 製造執行系統 (MES)

一種軟體系統，用於追蹤、監控、記錄和控制生產程序，將原物料轉換為現場成品。

## MAP

請參閱[遷移加速計劃](#)。

## 機制

建立工具、推動工具採用，然後檢查結果以進行調整的完整程序。機制是在操作時強化和改善自身的循環。如需詳細資訊，請參閱 AWS Well-Architected Framework 中的[建置機制](#)。

## 成員帳戶

除了屬於組織一部分的管理帳戶 AWS 帳戶 之外的所有 AWS Organizations。帳戶一次只能是一個組織的成員。

## 製造執行系統

請參閱[製造執行系統](#)。

## 訊息佇列遙測傳輸 (MQTT)

根據[發佈/訂閱](#)模式的輕量型 machine-to-machine (M2M) 通訊協定，適用於資源受限的 [IoT](#) 裝置。

## 微服務

一種小型的獨立服務，它可透過定義明確的 API 進行通訊，通常由小型獨立團隊擁有。例如，保險系統可能包含對應至業務能力 (例如銷售或行銷) 或子領域 (例如購買、索賠或分析) 的微服務。微服務的優點包括靈活性、彈性擴展、輕鬆部署、可重複使用的程式碼和適應力。如需詳細資訊，請參閱[使用無 AWS 伺服器服務整合微服務](#)。

## 微服務架構

一種使用獨立元件來建置應用程式的方法，這些元件會以微服務形式執行每個應用程式程序。這些微服務會使用輕量型 API，透過明確定義的介面進行通訊。此架構中的每個微服務都可以進行

更新、部署和擴展，以滿足應用程式特定功能的需求。如需詳細資訊，請參閱[在上實作微服務 AWS](#)。

## Migration Acceleration Program (MAP)

一種 AWS 計畫，提供諮詢支援、訓練和服務，協助組織建立強大的營運基礎，以移至雲端，並協助抵銷遷移的初始成本。MAP 包括用於有條不紊地執行舊式遷移的遷移方法以及一組用於自動化和加速常見遷移案例的工具。

### 大規模遷移

將大部分應用程式組合依波次移至雲端的程序，在每個波次中，都會以更快的速度移動更多應用程式。此階段使用從早期階段學到的最佳實務和經驗教訓來實作團隊、工具和流程的遷移工廠，以透過自動化和敏捷交付簡化工作負載的遷移。這是[AWS 遷移策略](#)的第三階段。

### 遷移工廠

可透過自動化、敏捷的方法簡化工作負載遷移的跨職能團隊。遷移工廠團隊通常包括營運、業務分析師和擁有者、遷移工程師、開發人員以及從事 Sprint 工作的 DevOps 專業人員。20% 至 50% 之間的企業應用程式組合包含可透過工廠方法優化的重複模式。如需詳細資訊，請參閱此內容集中的[遷移工廠的討論](#)和[雲端遷移工廠指南](#)。

### 遷移中繼資料

有關完成遷移所需的應用程式和伺服器的資訊。每種遷移模式都需要一組不同的遷移中繼資料。遷移中繼資料的範例包括目標子網路、安全群組和 AWS 帳戶。

### 遷移模式

可重複的遷移任務，詳細描述遷移策略、遷移目的地以及所使用的遷移應用程式或服務。範例：使用 AWS Application Migration Service 重新託管遷移至 Amazon EC2。

### 遷移組合評定 (MPA)

線上工具，提供驗證商業案例以遷移至的資訊 AWS 雲端。MPA 提供詳細的組合評定 (伺服器適當規模、定價、總體擁有成本比較、遷移成本分析) 以及遷移規劃 (應用程式資料分析和資料收集、應用程式分組、遷移優先順序，以及波次規劃)。[MPA 工具](#) (需要登入) 可供所有 AWS 顧問和 APN 合作夥伴顧問免費使用。

### 遷移準備程度評定 (MRA)

使用 AWS CAF 取得組織雲端整備狀態的洞見、識別優缺點，以及建立行動計劃以消除已識別差距的程序。如需詳細資訊，請參閱[遷移準備程度指南](#)。MRA 是[AWS 遷移策略](#)的第一階段。

## 遷移策略

用來將工作負載遷移至的方法 AWS 雲端。如需詳細資訊，請參閱本詞彙表中的 [7 個 Rs](#) 項目，並請參閱[動員您的組織以加速大規模遷移](#)。

## 機器學習 (ML)

請參閱[機器學習](#)。

## 現代化

將過時的 (舊版或單一) 應用程式及其基礎架構轉換為雲端中靈活、富有彈性且高度可用的系統，以降低成本、提高效率並充分利用創新。如需詳細資訊，請參閱 [《》中的現代化應用程式的策略 AWS 雲端](#)。

## 現代化準備程度評定

這項評估可協助判斷組織應用程式的現代化準備程度；識別優點、風險和相依性；並確定組織能夠在多大程度上支援這些應用程式的未來狀態。評定的結果就是目標架構的藍圖、詳細說明現代化程序的開發階段和里程碑的路線圖、以及解決已發現的差距之行動計畫。如需詳細資訊，請參閱 [《》中的評估應用程式的現代化準備 AWS 雲端](#) 程度。

## 單一應用程式 (單一)

透過緊密結合的程序作為單一服務執行的應用程式。單一應用程式有幾個缺點。如果一個應用程式功能遇到需求激增，則必須擴展整個架構。當程式碼庫增長時，新增或改進單一應用程式的功能也會變得更加複雜。若要解決這些問題，可以使用微服務架構。如需詳細資訊，請參閱[將單一體系分解為微服務](#)。

## MPA

請參閱[遷移產品組合評估](#)。

## MQTT

請參閱[訊息佇列遙測傳輸](#)。

## 多類別分類

一個有助於產生多類別預測的過程 (預測兩個以上的結果之一)。例如，機器學習模型可能會詢問「此產品是書籍、汽車還是電話？」或者「這個客戶對哪種產品類別最感興趣？」

## 可變基礎設施

更新和修改生產工作負載現有基礎設施的模型。為了提高一致性、可靠性和可預測性，AWS Well-Architected Framework 建議使用[不可變基礎設施](#)做為最佳實務。

## O

### OAC

請參閱[原始存取控制](#)。

### OAI

請參閱[原始存取身分](#)。

### OCM

請參閱[組織變更管理](#)。

### 離線遷移

一種遷移方法，可在遷移過程中刪除來源工作負載。此方法涉及延長停機時間，通常用於小型非關鍵工作負載。

### OI

請參閱[操作整合](#)。

### OLA

請參閱[操作層級協議](#)。

### 線上遷移

一種遷移方法，無需離線即可將來源工作負載複製到目標系統。連接至工作負載的應用程式可在遷移期間繼續運作。此方法涉及零至最短停機時間，通常用於關鍵的生產工作負載。

### OPC-UA

請參閱[開放程序通訊 - 統一架構](#)。

### 開放程序通訊 - 統一架構 (OPC-UA)

用於工業自動化machine-to-machine(M2M) 通訊協定。OPC-UA 提供資料加密、身分驗證和授權機制的互通性標準。

### 操作水準協議 (OLA)

一份協議，闡明 IT 職能群組承諾向彼此提供的內容，以支援服務水準協議 (SLA)。

### 操作整備審查 (ORR)

問題和相關最佳實務的檢查清單，可協助您了解、評估、預防或減少事件和可能失敗的範圍。如需詳細資訊，請參閱 AWS Well-Architected Framework 中的[操作準備度審查 \(ORR\)](#)。

## 操作技術 (OT)

使用實體環境控制工業操作、設備和基礎設施的硬體和軟體系統。在製造中，OT 和資訊技術 (IT) 系統的整合是 [工業 4.0](#) 轉型的關鍵重點。

## 操作整合 (OI)

在雲端中將操作現代化的程序，其中包括準備程度規劃、自動化和整合。如需詳細資訊，請參閱 [操作整合指南](#)。

## 組織追蹤

由建立的線索 AWS CloudTrail 會記錄 AWS 帳戶組織中所有的所有事件 AWS Organizations。在屬於組織的每個 AWS 帳戶中建立此追蹤，它會跟蹤每個帳戶中的活動。如需詳細資訊，請參閱 CloudTrail 文件中的 [建立組織追蹤](#)。

## 組織變更管理 (OCM)

用於從人員、文化和領導力層面管理重大、顛覆性業務轉型的架構。OCM 透過加速變更採用、解決過渡問題，以及推動文化和組織變更，協助組織為新系統和策略做好準備，並轉移至新系統和策略。在 AWS 遷移策略中，此架構稱為人員加速，因為雲端採用專案所需的變更速度。如需詳細資訊，請參閱 [OCM 指南](#)。

## 原始存取控制 (OAC)

CloudFront 中的增強型選項，用於限制存取以保護 Amazon Simple Storage Service (Amazon S3) 內容。OAC 支援使用 S3 AWS KMS (SSE-KMS) 的所有伺服器端加密中的所有 S3 儲存貯體 AWS 區域，以及對 S3 儲存貯體的動態 PUT 和 DELETE 請求。

## 原始存取身分 (OAI)

CloudFront 中的一個選項，用於限制存取以保護 Amazon S3 內容。當您使用 OAI 時，CloudFront 會建立一個可供 Amazon S3 進行驗證的主體。經驗證的主體只能透過特定 CloudFront 分發來存取 S3 儲存貯體中的內容。另請參閱 [OAC](#)，它可提供更精細且增強的存取控制。

## ORR

請參閱 [操作整備審核](#)。

## OT

請參閱 [操作技術](#)。

## 傳出 (輸出) VPC

在 AWS 多帳戶架構中，處理從應用程式內啟動之網路連線的 VPC。[AWS 安全參考架構](#)建議您使用傳入、傳出和檢查 VPC 來設定網路帳戶，以保護應用程式與更廣泛的網際網路之間的雙向介面。

## P

### 許可界限

附接至 IAM 主體的 IAM 管理政策，可設定使用者或角色擁有的最大許可。如需詳細資訊，請參閱 IAM 文件中的[許可界限](#)。

### 個人身分識別資訊 (PII)

直接檢視或與其他相關資料配對時，可用來合理推斷個人身分的資訊。PII 的範例包括名稱、地址和聯絡資訊。

### PII

請參閱[個人身分識別資訊](#)。

### 手冊

一組預先定義的步驟，可擷取與遷移關聯的工作，例如在雲端中提供核心操作功能。手冊可以採用指令碼、自動化執行手冊或操作現代化環境所需的程序或步驟摘要的形式。

### PLC

請參閱[可程式設計邏輯控制器](#)。

### PLM

請參閱[產品生命週期管理](#)。

### 政策

可定義許可的物件（請參閱[身分型政策](#)）、指定存取條件（請參閱[資源型政策](#)），或定義組織中所有帳戶的最大許可 AWS Organizations（請參閱[服務控制政策](#)）。

### 混合持久性

根據資料存取模式和其他需求，獨立選擇微服務的資料儲存技術。如果您的微服務具有相同的資料儲存技術，則其可能會遇到實作挑戰或效能不佳。如果微服務使用最適合其需求的資料儲存，則可以更輕鬆地實作並達到更好的效能和可擴展性。

## 組合評定

探索、分析應用程式組合並排定其優先順序以規劃遷移的程序。如需詳細資訊，請參閱[評估遷移準備程度](#)。

## 述詞

傳回 true 或的查詢條件 false，通常位於 WHERE 子句中。

## 述詞下推

一種資料庫查詢最佳化技術，可在傳輸前篩選查詢中的資料。這可減少必須從關聯式資料庫擷取和處理的資料量，並改善查詢效能。

## 預防性控制

旨在防止事件發生的安全控制。這些控制是第一道防線，可協助防止對網路的未經授權存取或不必要變更。如需詳細資訊，請參閱在 AWS 上實作安全控制中的[預防性控制](#)。

## 委託人

中可執行動作和存取資源 AWS 的實體。此實體通常是 AWS 帳戶、IAM 角色或使用者的根使用者。如需詳細資訊，請參閱 IAM 文件中[角色術語和概念](#)中的主體。

## 依設計的隱私權

透過整個開發程序將隱私權納入考量的系統工程方法。

## 私有託管區域

一種容器，它包含有關您希望 Amazon Route 53 如何回應一個或多個 VPC 內的域及其子域之 DNS 查詢的資訊。如需詳細資訊，請參閱 Route 53 文件中的[使用私有託管區域](#)。

## 主動控制

旨在防止部署不合規資源的[安全控制](#)。這些控制項會在佈建資源之前對其進行掃描。如果資源不符合控制項，則不會佈建。如需詳細資訊，請參閱 AWS Control Tower 文件中的[控制項參考指南](#)，並參閱實作安全[控制項中的主動](#)控制項。 AWS

## 產品生命週期管理 (PLM)

管理產品整個生命週期的資料和程序，從設計、開發和啟動，到成長和成熟，再到拒絕和移除。

## 生產環境

請參閱[環境](#)。

## 可程式設計邏輯控制器 (PLC)

在製造中，高度可靠、可調整的電腦，可監控機器並自動化製造程序。

### 提示鏈結

使用一個 [LLM](#) 提示的輸出做為下一個提示的輸入，以產生更好的回應。此技術用於將複雜任務分解為子任務，或反覆精簡或展開初步回應。它有助於提高模型回應的準確性和相關性，並允許更精細、個人化的結果。

### 擬匿名化

將資料集中的個人識別符取代為預留位置值的程序。假名化有助於保護個人隱私權。假名化資料仍被視為個人資料。

### 發佈/訂閱 (pub/sub)

一種模式，可啟用微服務之間的非同步通訊，以提高可擴展性和回應能力。例如，在微服務型 [MES](#) 中，微服務可以將事件訊息發佈到其他微服務可訂閱的頻道。系統可以新增新的微服務，而無需變更發佈服務。

## Q

### 查詢計劃

一系列步驟，如指示，用於存取 SQL 關聯式資料庫系統中的資料。

### 查詢計劃迴歸

在資料庫服務優化工具選擇的計畫比對資料庫環境進行指定的變更之前的計畫不太理想時。這可能因為對統計資料、限制條件、環境設定、查詢參數繫結的變更以及資料庫引擎的更新所導致。

## R

### RACI 矩陣

請參閱 [負責、負責、諮詢、告知 \(RACI\)](#)。

### RAG

請參閱 [擷取增強生成](#)。

## 勒索軟體

一種惡意軟體，旨在阻止對計算機系統或資料的存取，直到付款為止。

## RASCI 矩陣

請參閱[負責、負責、諮詢、告知 \(RACI\)](#)。

## RCAC

請參閱[資料列和資料欄存取控制](#)。

## 僅供讀取複本

用於唯讀用途的資料庫複本。您可以將查詢路由至僅供讀取複本以減少主資料庫的負載。

## 重新架構師

請參閱[7 個 R](#)。

## 復原點目標 (RPO)

自上次資料復原點以來可接受的時間上限。這會決定最後一個復原點與服務中斷之間可接受的資料遺失。

## 復原時間目標 (RTO)

服務中斷與服務還原之間的可接受延遲上限。

## 重構

請參閱[7 個 R](#)。

## 區域

地理區域中的 AWS 資源集合。每個 AWS 區域 都獨立於其他，以提供容錯能力、穩定性和彈性。如需詳細資訊，請參閱[指定 AWS 區域 您的帳戶可以使用哪些](#)。

## 迴歸

預測數值的 ML 技術。例如，為了解決「這房子會賣什麼價格？」的問題 ML 模型可以使用線性迴歸模型，根據已知的房屋事實 (例如，平方英尺) 來預測房屋的銷售價格。

## 重新託管

請參閱[7 個 R](#)。

## 版本

在部署程序中，它是將變更提升至生產環境的動作。

## 重新放置

請參閱 [7 個 R](#)。

## Replatform

請參閱 [7 個 R](#)。

## 回購

請參閱 [7 個 R](#)。

## 彈性

應用程式抵禦中斷或從中斷中復原的能力。[在中規劃彈性時，高可用性和災難復原](#)是常見的考量 AWS 雲端。如需詳細資訊，請參閱[AWS 雲端 彈性](#)。

## 資源型政策

附接至資源的政策，例如 Amazon S3 儲存貯體、端點或加密金鑰。這種類型的政策會指定允許存取哪些主體、支援的動作以及必須滿足的任何其他條件。

## 負責者、當責者、事先諮詢者和事後告知者 (RACI) 矩陣

矩陣，定義所有參與遷移活動和雲端操作之各方的角色和責任。矩陣名稱衍生自矩陣中定義的責任類型：負責人 (R)、責任 (A)、已諮詢 (C) 和知情 (I)。支援 (S) 類型為選用。如果您包含支援，則矩陣稱為 RASCI 矩陣，如果您排除它，則稱為 RACI 矩陣。

## 回應性控制

一種安全控制，旨在驅動不良事件或偏離安全基準的補救措施。如需詳細資訊，請參閱在 AWS 上實作安全控制中的[回應性控制](#)。

## 保留

請參閱 [7 個 R](#)。

## 淘汰

請參閱 [7 個 R](#)。

## 檢索增強生成 (RAG)

[一種生成式 AI](#) 技術，其中 [LLM](#) 會在產生回應之前參考訓練資料來源以外的授權資料來源。例如，RAG 模型可能會對組織的知識庫或自訂資料執行語意搜尋。如需詳細資訊，請參閱[什麼是 RAG](#)。

## 輪換

定期更新[秘密](#)的程序，讓攻擊者更難存取登入資料。

## 資料列和資料欄存取控制 (RCAC)

使用已定義存取規則的基本、彈性 SQL 表達式。RCAC 包含資料列許可和資料欄遮罩。

## RPO

請參閱[復原點目標](#)。

## RTO

請參閱[復原時間目標](#)。

## 執行手冊

執行特定任務所需的一組手動或自動程序。這些通常是為了簡化重複性操作或錯誤率較高的程序而建置。

# S

## SAML 2.0

許多身分提供者 (IdP) 使用的開放標準。此功能會啟用聯合單一登入 (SSO)，AWS 管理主控台 讓使用者可以登入 或呼叫 AWS API 操作，而不必為組織中的每個人在 IAM 中建立使用者。如需有關以 SAML 2.0 為基礎的聯合詳細資訊，請參閱 IAM 文件中的[關於以 SAML 2.0 為基礎的聯合](#)。

## 斯卡達

請參閱[監督控制和資料擷取](#)。

## SCP

請參閱[服務控制政策](#)。

## 秘密

以加密形式存放的 AWS Secrets Manager 機密或限制資訊，例如密碼或使用者登入資料。它由秘密值及其中繼資料組成。秘密值可以是二進位、單一字串或多個字串。如需詳細資訊，請參閱[Secrets Manager 秘密中的內容？](#) Secrets Manager 文件中的。

## 依設計的安全性

透過整個開發程序將安全性納入考量的系統工程方法。

## 安全控制

一種技術或管理防護機制，它可預防、偵測或降低威脅行為者利用安全漏洞的能力。安全控制有四種主要類型：[預防性](#)、[偵測性](#)、[回應性](#)和[主動性](#)。

## 安全強化

減少受攻擊面以使其更能抵抗攻擊的過程。這可能包括一些動作，例如移除不再需要的資源、實作授予最低權限的安全最佳實務、或停用組態檔案中不必要的功能。

### 安全資訊與事件管理 (SIEM) 系統

結合安全資訊管理 (SIM) 和安全事件管理 (SEM) 系統的工具與服務。SIEM 系統會收集、監控和分析來自伺服器、網路、裝置和其他來源的資料，以偵測威脅和安全漏洞，並產生提醒。

### 安全回應自動化

預先定義和程式設計的動作，旨在自動回應或修復安全事件。這些自動化可做為[偵測或回應](#)式安全控制，協助您實作 AWS 安全最佳實務。自動化回應動作的範例包括修改 VPC 安全群組、修補 Amazon EC2 執行個體或輪換登入資料。

### 伺服器端加密

由接收資料的 AWS 服務 在其目的地加密資料。

### 服務控制政策 (SCP)

為 AWS Organizations 中的組織的所有帳戶提供集中控制許可的政策。SCP 會定義防護機制或設定管理員可委派給使用者或角色的動作限制。您可以使用 SCP 作為允許清單或拒絕清單，以指定允許或禁止哪些服務或動作。如需詳細資訊，請參閱 AWS Organizations 文件中的[服務控制政策](#)。

### 服務端點

的進入點 URL AWS 服務。您可以使用端點，透過程式設計方式連接至目標服務。如需詳細資訊，請參閱 AWS 一般參考 中的 [AWS 服務 端點](#)。

### 服務水準協議 (SLA)

一份協議，闡明 IT 團隊承諾向客戶提供的服務，例如服務正常執行時間和效能。

### 服務層級指標 (SLI)

服務效能方面的測量，例如其錯誤率、可用性或輸送量。

### 服務層級目標 (SLO)

代表服務運作狀態的目標指標，由[服務層級指標](#)測量。

### 共同責任模式

描述您與共同 AWS 承擔雲端安全與合規責任的模型。AWS 負責雲端的安全，而負責雲端的安全。如需詳細資訊，請參閱[共同責任模式](#)。

## SIEM

請參閱[安全資訊和事件管理系統](#)。

## 單一故障點 (SPOF)

應用程式的單一關鍵元件故障，可能會中斷系統。

## SLA

請參閱[服務層級協議](#)。

## SLI

請參閱[服務層級指標](#)。

## SLO

請參閱[服務層級目標](#)。

## 先拆分後播種模型

擴展和加速現代化專案的模式。定義新功能和產品版本時，核心團隊會進行拆分以建立新的產品團隊。這有助於擴展組織的能力和服務，提高開發人員生產力，並支援快速創新。如需詳細資訊，請參閱[中的階段式應用程式現代化方法 AWS 雲端](#)。

## SPOF

請參閱[單一故障點](#)。

## 星狀結構描述

使用一個大型事實資料表來存放交易或測量資料的資料庫組織結構，並使用一或多個較小的維度資料表來存放資料屬性。此結構旨在用於[資料倉儲](#)或商業智慧用途。

## Strangler Fig 模式

一種現代化單一系統的方法，它會逐步重寫和取代系統功能，直到舊式系統停止使用為止。此模式源自無花果藤，它長成一棵馴化樹並最終戰勝且取代了其宿主。該模式由[Martin Fowler 引入](#)，作為重寫單一系統時管理風險的方式。如需有關如何套用此模式的範例，請參閱[使用容器和 Amazon API Gateway 逐步現代化舊版 Microsoft ASP.NET \(ASMX\) Web 服務](#)。

## 子網

您 VPC 中的 IP 地址範圍。子網必須位於單一可用區域。

## 監控控制和資料擷取 (SCADA)

在製造中，使用硬體和軟體來監控實體資產和生產操作的系統。

## 對稱加密

使用相同金鑰來加密及解密資料的加密演算法。

## 合成測試

以模擬使用者互動的方式測試系統，以偵測潛在問題或監控效能。您可以使用 [Amazon CloudWatch Synthetics](#) 來建立這些測試。

## 系統提示

一種向 [LLM](#) 提供內容、指示或指導方針以指示其行為的技術。系統提示有助於設定內容，並建立與使用者互動的規則。

# T

## 標籤

做為中繼資料以組織 AWS 資源的鍵值對。標籤可協助您管理、識別、組織、搜尋及篩選資源。如需詳細資訊，請參閱 [標記您的 AWS 資源](#)。

## 目標變數

您嘗試在受監督的 ML 中預測的值。這也被稱為結果變數。例如，在製造設定中，目標變數可能是產品瑕疵。

## 任務清單

用於透過執行手冊追蹤進度的工具。任務清單包含執行手冊的概觀以及要完成的一般任務清單。對於每個一般任務，它包括所需的預估時間量、擁有者和進度。

## 測試環境

請參閱 [環境](#)。

## 訓練

為 ML 模型提供資料以供學習。訓練資料必須包含正確答案。學習演算法會在訓練資料中尋找將輸入資料屬性映射至目標的模式 (您想要預測的答案)。它會輸出擷取這些模式的 ML 模型。可以使用 ML 模型，來預測您不知道的目標新資料。

## 傳輸閘道

可以用於互連 VPC 和內部部署網路的網路傳輸中樞。如需詳細資訊，請參閱 AWS Transit Gateway 文件中的 [什麼是傳輸閘道](#)。

## 主幹型工作流程

這是一種方法，開發人員可在功能分支中本地建置和測試功能，然後將這些變更合併到主要分支中。然後，主要分支會依序建置到開發環境、生產前環境和生產環境中。

## 受信任的存取權

將許可授予您指定的服務，以代表您在組織中 AWS Organizations 及其帳戶中執行任務。受信任的服務會在需要該角色時，在每個帳戶中建立服務連結角色，以便為您執行管理工作。如需詳細資訊，請參閱文件中的 AWS Organizations [搭配使用 AWS Organizations 與其他 AWS 服務](#)。

## 調校

變更訓練程序各個層面，以提高 ML 模型的準確性。例如，可以透過產生標籤集、新增標籤、然後在不同的設定下多次重複這些步驟來訓練 ML 模型，以優化模型。

## 雙比薩團隊

兩個比薩就能吃飽的小型 DevOps 團隊。雙披薩團隊規模可確保軟體開發中的最佳協作。

# U

## 不確定性

這是一個概念，指的是不精確、不完整或未知的資訊，其可能會破壞預測性 ML 模型的可靠性。有兩種類型的不確定性：認知不確定性是由有限的、不完整的資料引起的，而隨機不確定性是由資料中固有的噪聲和隨機性引起的。

## 未區分的任務

也稱為繁重工作，是建立和操作應用程式的必要工作，但不為最終使用者提供直接價值或提供競爭優勢。未區分任務的範例包括採購、維護和容量規劃。

## 較高的環境

請參閱 [環境](#)。

# V

## 清空

一種資料庫維護操作，涉及增量更新後的清理工作，以回收儲存並提升效能。

## 版本控制

追蹤變更的程序和工具，例如儲存庫中原始程式碼的變更。

## VPC 對等互連

兩個 VPC 之間的連線，可讓您使用私有 IP 地址路由流量。如需詳細資訊，請參閱 Amazon VPC 文件中的[什麼是 VPC 對等互連](#)。

## 漏洞

危及系統安全性的軟體或硬體瑕疵。

# W

## 暖快取

包含經常存取的目前相關資料的緩衝快取。資料庫執行個體可以從緩衝快取讀取，這比從主記憶體或磁碟讀取更快。

## 暖資料

不常存取的資料。查詢這類資料時，通常可接受中等速度的查詢。

## 視窗函數

SQL 函數，對與目前記錄在某種程度上相關的資料列群組執行計算。視窗函數適用於處理任務，例如根據目前資料列的相對位置計算移動平均值或存取資料列的值。

## 工作負載

提供商業價值的資源和程式碼集合，例如面向客戶的應用程式或後端流程。

## 工作串流

遷移專案中負責一組特定任務的功能群組。每個工作串流都是獨立的，但支援專案中的其他工作串流。例如，組合工作串流負責排定應用程式、波次規劃和收集遷移中繼資料的優先順序。組合工作串流將這些資產交付至遷移工作串流，然後再遷移伺服器 and 應用程式。

## WORM

請參閱[寫入一次，讀取許多](#)。

## WQF

請參閱[AWS 工作負載資格架構](#)。

## 寫入一次，讀取許多 (WORM)

儲存模型，可一次性寫入資料，並防止刪除或修改資料。授權使用者可以視需要多次讀取資料，但無法變更資料。此資料儲存基礎設施被視為[不可變](#)。

## Z

### 零時差入侵

利用[零時差漏洞](#)的攻擊，通常是惡意軟體。

### 零時差漏洞

生產系統中未緩解的瑕疵或漏洞。威脅行為者可以使用這種類型的漏洞來攻擊系統。開發人員經常因為攻擊而意識到漏洞。

### 零鏡頭提示

提供 [LLM](#) 執行任務的指示，但沒有可協助引導任務的範例 (快照)。LLM 必須使用其預先訓練的知識來處理任務。零鏡頭提示的有效性取決於任務的複雜性和提示的品質。另請參閱[少量擷取提示](#)。

### 殭屍應用程式

CPU 和記憶體平均使用率低於 5% 的應用程式。在遷移專案中，通常會淘汰這些應用程式。

本文為英文版的機器翻譯版本，如內容有任何歧義或不一致之處，概以英文版為準。