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

AWS 方案指引

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

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
..... xxxviii

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，以及摘要記錄。

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以下生成式 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 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 SDKsAmazon 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，擷取醫療文字並為醫療 NLP 任務提供結果。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 為同步實體偵測提供下列文字分析 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 知識庫中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 Systematized Nomenclature of Medicine， clinical Terms (SNOMED CT) 版本的程式碼。SNOMED CT 提供全方位的醫療概念詞彙，包括醫療情況和解剖，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內容資訊，例如特徵和屬性。然後，SNOMED CT 概念 IDs 可用於建構患者臨床資料，以便與 SNOMED CT 多階層搭配使用進行醫療編碼、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
- 偵測受保護醫療資訊 (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。由於這無法透過 Amazon Comprehend Medical 取得，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：

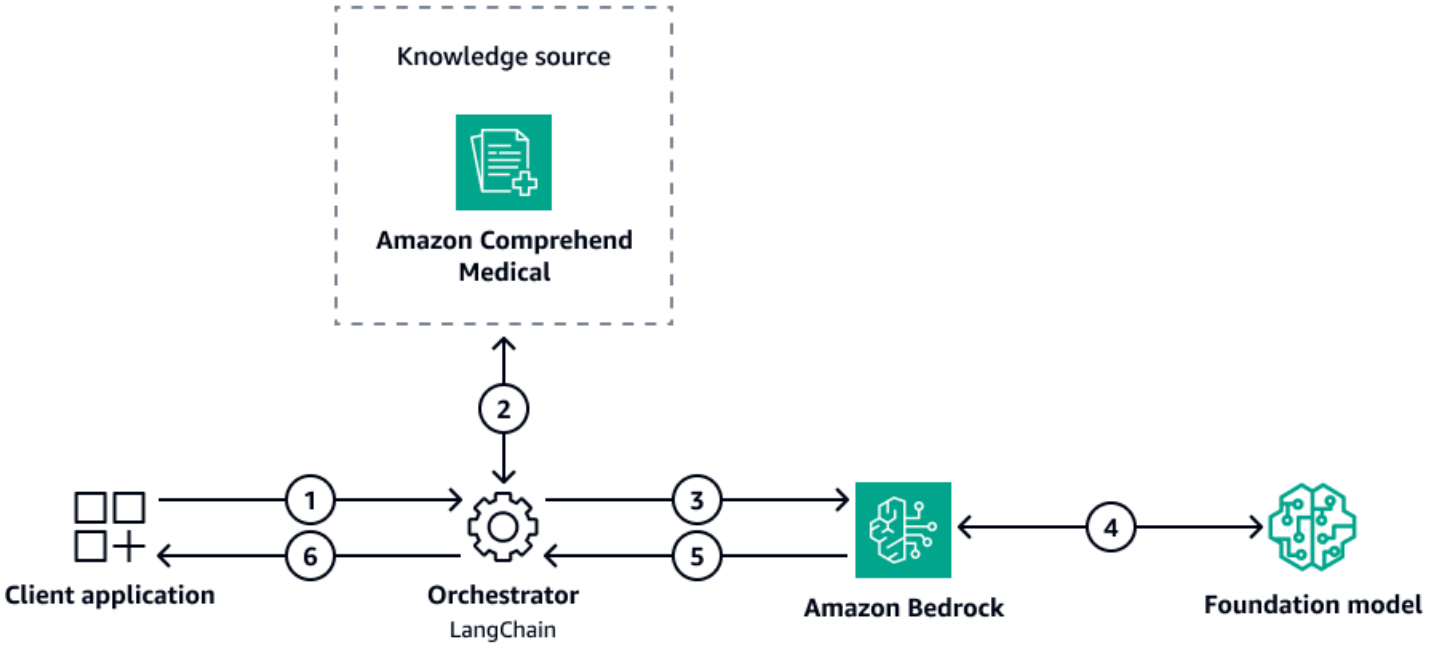
- 使用 Amazon Comprehend Medical 的初始結果作為 LLM 的內容，以提高實體選擇的準確性
- 實作自訂實體辨識、摘要、問答和其他使用案例

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中識別 ICD-10-CM 診斷碼的 RAG 方法。它使用 Amazon Comprehend Medical 做為知識來源。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：

1. 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。該患者昨天晚上前往 緊急室，有大約 7 天到 8 天持續的腹痛史。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。」
2. 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，將擷取的醫療資訊與up-to-date研究和指導方針結合在一起

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 – 製作 LLM 提示詞時，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。
- 提供特定網域的知識 – 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。考慮使用其他領域特定的知識來源來補充 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、安全和醫療保健中 AI 系統負責任使用相關的準則。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 提供者，請遵循該提供者的建議。

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

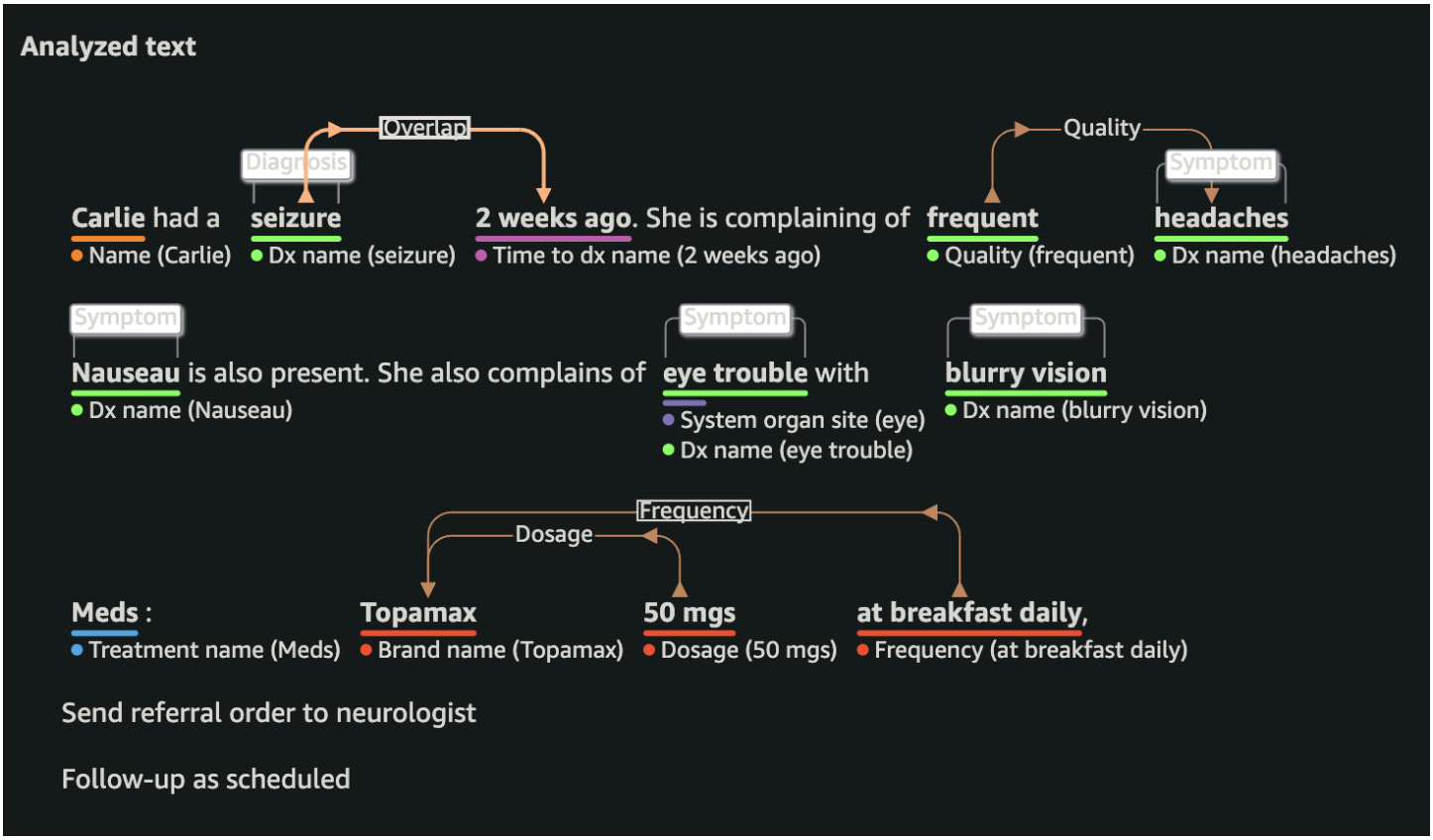
Amazon Comprehend Medical 通常會提供大量資訊。您可能想要減少醫療專業人員必須檢閱的結果數量。在這種情況下，您可以使用 LLM 來篩選這些結果。Amazon Comprehend Medical 實體包含可信度分數，您可以在設計提示時用作篩選機制。

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 代碼，以因應癱瘓和頭煩。

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、非難治、無狀態癱瘓	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嘆號和驚嘆號，沒有狀態驚嘆號	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，會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 freqeunt 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>
        <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，使用這些 LLM 時（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專門醫療保健任務。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，此方法在處理模型訓練資料之外的高度專業醫療知識時，或需要絕對精確度以進行重大照護決策時，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。

以下是兩個常見的使用案例，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之醫學知識的 corpus，例如 文件。這是根據相似性或情緒使用一對文字來完成，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進行訓練的自訂 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
BioBERT	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
ClinicalBERT	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
ClinicalGPT	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、特定領域的知識和多回合對話諮詢	解碼器
GatorTron-OG	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
Med-BERT	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
Med-PaLM	用於醫療用途的問題答案	醫療和生物醫學文字的資料集	解碼器
medAlpaca	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
BiomedBERT	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。解碼器適用於產生任務。
- 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，我們建議微調現有的模型。

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。

本節根據以下兩個醫療保健客戶的高影響使用案例進行 AWS 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程記錄，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大型敏感資料集的訓練成本，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同運作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的大小而增加，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) 會教導模型使用結構化輸入輸出範例來遵循明確指示。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來最佳化模型行為。使用根據人力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 非常適合讓輸出符合臨床準則，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。這可讓模型以易於使用的語言傳達複雜的診斷，同時仍然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：

- 根據微調方法的磁碟區
- 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
網域調整預先訓練	超過 100，000 個網域文字
監督微調	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，應該是資料策劃過程的一部分，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機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 文件中的支援的模型和區域，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 APIs Amazon Bedrock。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 等開放原始碼模型可在 Amazon SageMaker AI 服務間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，這些模型是理想的選擇。

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，提供一些靈活性並降低複雜性。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，但最適合在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。

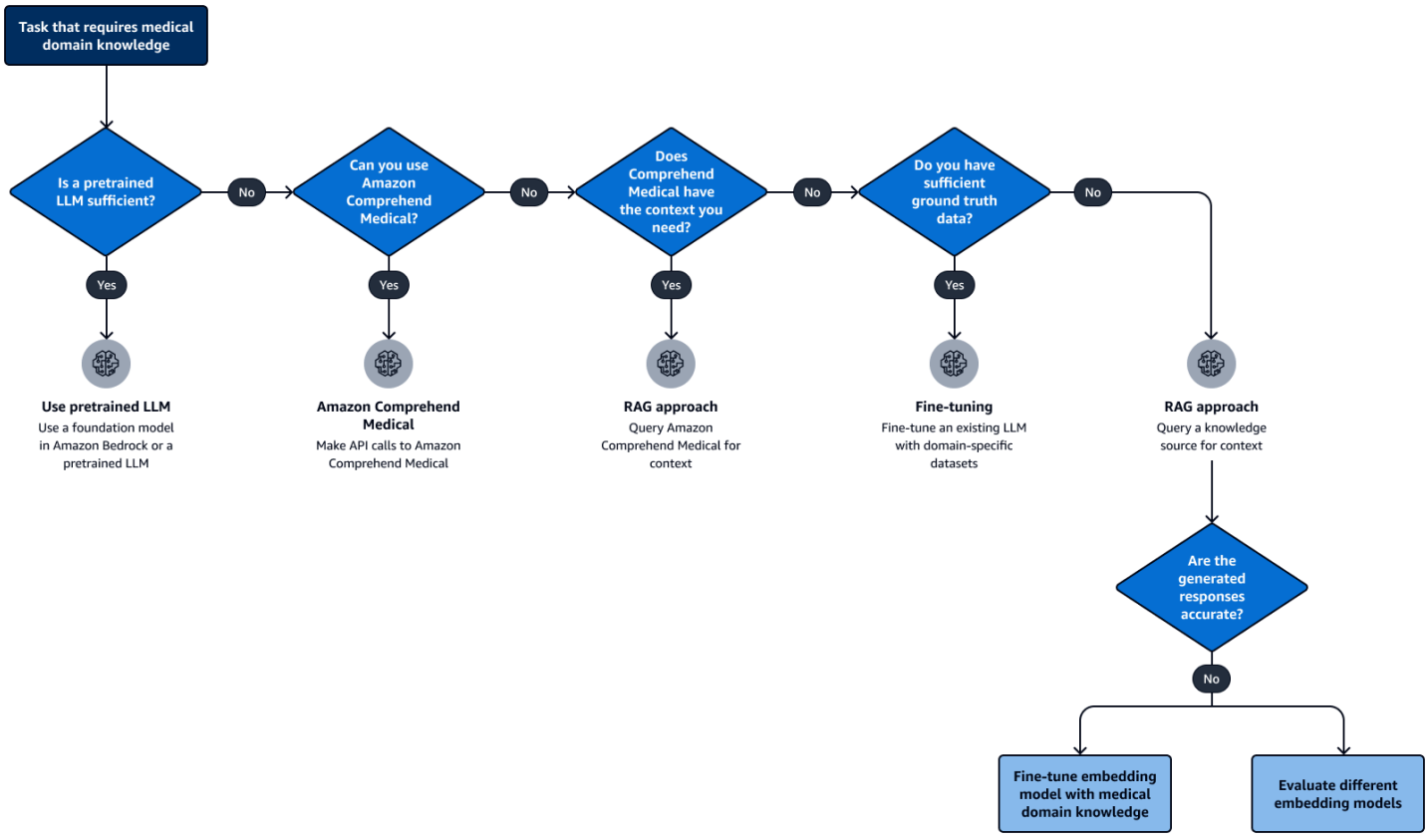
MLflow 是一種開放原始碼解決方案，可協助您追蹤微調實驗。Amazon SageMaker AI 原生支援 MLflow，可協助您以視覺化方式比較訓練執行中的指標。對於 Amazon Bedrock 上的 微調任務，指標會串流到 Amazon CloudWatch，讓您可以在 CloudWatch 主控台中視覺化指標。

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 本節說明下列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 時面臨不同的複雜性層級，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。您可能需要聘請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、成本較低的方法一致地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中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
新區段	我們在醫療保健章節和提示工程章節中新增了微調大型語言模型。 https://docs.aws.amazon.com/prescriptive-guidance/latest/generative-ai-nlp-healthcare/llms.html#llm-customization-prompt-engineering	2025 年 12 月 5 日
初次出版	—	2024 年 12 月 16 日

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。