一、報名內容填寫

1. **團隊介紹**

我們是「序識人 BioSeer」，一個由資訊工程與生物科技背景組成的跨領域創新團隊。團隊成員皆對人工智慧於生物醫學領域的應用充滿熱忱，並具備從基礎研究到模型實作的完整實戰經驗。團隊名稱「序識人」象徵我們擁有解析生物序列（Sequence）智慧（Sense）的能力，也暗含「救世人」之意，呼應我們期望運用科技改善人類健康的初衷。

我們的核心技術團隊成員曾參與多場AI與智慧醫療主題競賽，具備深度學習模型建構、蛋白質序列分析與視覺化界面設計能力。在一次智慧物聯網課程專題中，我們實作了首個蛋白質功能預測模型雛型，後續更進一步導入先進的蛋白質語言模型（ESM2），並搭配多視窗卷積神經網路（MCNN）進行結構優化，使整體準確率顯著提升。

除了技術研發，我們也關注模型的實際應用潛力。我們團隊定期與生命科學背景的教授與業界生技顧問進行交流，確保技術導入符合現場需求。在設計上，我們特別重視模型可視化與使用者體驗，期望最終可將此系統發展為健康大數據分析平台中的一個重要模組，為智慧醫療提供可靠而高效的蛋白質功能預測服務。

1. **目標客群與痛點說明**

在智慧醫療與健康大數據發展的時代，醫療單位與研究機構每日都面臨大量蛋白質序列資料，但由於這些序列缺乏結構資訊與功能註解，無法立即投入分析使用，嚴重影響資料運用效率。尤其是在疾病機轉研究、藥物靶點發掘、生物標誌物預測等應用中，若能快速準確地辨識出特定功能蛋白，如電子傳遞蛋白（ET）與轉運蛋白（TP），將大幅縮短研發時程與成本。

然而，目前主流的蛋白質功能註解工具（如 BLAST、Pfam）多依賴序列比對，面對低相似度蛋白表現不佳，且流程冗長，對於需要高通量處理的情境難以應對。此外，人工標註也存在主觀判斷與時間成本，無法滿足快速成長的蛋白質資料量。

我們的預測系統即是為了解決上述痛點而生。透過 ESM2 語言模型自動擷取蛋白序列中的高層次語意特徵，搭配 MCNN 架構分析不同尺度的結構片段，使系統具備從序列推斷功能的能力，不需人工標註或外部資料比對，即可快速分類未知蛋白為 ET 或 TP。

目標客群包含：

* 生技公司與藥廠：需大量分析蛋白功能作為藥物研發依據
* 醫學研究單位與醫院實驗室：需判讀患者樣本蛋白質組數據
* 生物資訊平台開發商：可將本模型模組化整合至既有系統

我們相信，這項技術不僅可大幅減少人力成本，也能提升資料使用效率，是智慧醫療資料處理流程中極具潛力的創新解方。

1. **AI技術運用說明**

本預測系統融合當前最先進的兩項人工智慧技術：語言模型與卷積神經網路。首先，採用 Meta AI 所發表的蛋白質語言模型 **ESM2 (Evolutionary Scale Modeling)**，將原始蛋白質序列轉換為上下文語意嵌入向量，使模型能捕捉到序列中隱含的生物學意涵。這種表徵方式突破了傳統手工特徵的限制，能夠適用於未知蛋白與低相似序列。

其次，我們設計了一個 **多視窗卷積神經網路（Multi-Channel CNN）** 模型，模仿生物學中蛋白結構多尺度的特性，讓模型同時從短片段與長片段中提取結構訊號。此設計使模型能更細緻地辨別出 ET 與 TP 之間潛在的序列差異。

模型訓練過程中，我們導入 Dropout、Early Stopping、Batch Normalization 等常見策略提升泛化能力，並以五折交叉驗證評估性能。在實驗中，模型在訓練與驗證集上皆展現穩定的預測準確率，平均高達 **92.6%**，F1-Score 與 AUC 亦顯示優異的分類效果。

技術實作上，我們使用 Python + PyTorch 框架進行開發，並於 GPU 環境中加速訓練。整體架構具備模組化潛力，未來可整合為 API，部署於雲端平台，便於第三方研究人員或企業串接應用。

1. **獨特競爭優勢**

「序識人 BioSeer」的預測系統具有明確的三大創新特色：

**1. 結合語言模型與多視窗卷積的架構創新**：傳統蛋白質分類方法多依賴比對工具或單一視窗特徵，而我們融合 ESM2 的 contextual embedding 與 MCNN 多尺度學習能力，打造出兼具深度與彈性的預測架構，在多樣性蛋白中仍具高度辨識力。

**2. 聚焦於 ET/TP 功能類別的稀有任務**：目前少有研究專門針對電子傳遞蛋白與轉運蛋白進行 AI 分類，然而這兩類蛋白在疾病研究與代謝通路中扮演關鍵角色。我們模型填補此領域的工具缺口，提供具有即戰力的解決方案。

**3. 高準確率＋可解釋性導向設計**：預計後續結合 Attention 機制與視覺化呈現，讓使用者可看見模型決策依據，提升使用信任度與科研應用價值。

此外，我們的架構具備高度擴充性，未來可應用至其他分類任務，如膜蛋白、酵素、結構域等，逐步建構出一套完整的 AI 蛋白質功能預測平台。我們相信，這項技術的原創力與彈性，將成為未來智慧醫療的基礎建設之一。

1. **預期效益**

此系統預期將在以下三個面向創造顯著價值：

⟡ 學術與研究應用價值：可大幅簡化蛋白質序列註解流程，協助研究人員即時聚焦分析目標，加速疾病機轉研究與新蛋白功能探索。

⟡ 商業模式與技術授權潛力：系統可發展為獨立模組整合至各類生技資料處理平台中，透過 API 提供即時預測服務。未來亦可考慮 SaaS 授權、按需收費，建立可持續營運模式。

⟡ 社會與醫療效益：透過準確預測關鍵蛋白功能，能有效支援新藥研發、生物標誌物驗證與疾病診斷工具開發，間接提升整體醫療品質與精準醫療發展。

長遠而言，這套系統不僅是一個技術成果，更代表我們期望以 AI 賦能生物醫學、加速資料價值釋放、實現智慧醫療願景的具體實踐。

二、pdf檔案規劃

🟦 頁 1：封面

* 背景可以設計成漸層（淺藍～深藍）或抽象幾何視覺
* 標題字大而醒目（36~48pt）
* 副標題（例如團隊名、領域、口號）適中（18~24pt）
* 建議加一張象徵性的視覺圖（例：蛋白質結構圖＋AI晶片＋城市線條混合）

🟩 頁 2：目標與痛點

* 左側放簡單情境插畫或痛點示意圖
* 右側以條列方式列出目標客群與具體痛點
* 用 icon + 關鍵詞輔助說明（例如使用 [🧬][🏥][🧪] 圖示）

🟧 頁 3：解決方案總覽

* 上半部是流程圖（Data → Feature → MCNN + ESM2 → Predict）
* 下半部用簡單句子說明「為何我們的解法與眾不同」
* 可以用時間軸或流程框強化技術邏輯感

🟥 頁 4：模型架構詳解

* 模型 Block Diagram 建議橫向放置，標註每個模塊（如 Input, Embedding, CNN Layers, FC Layer）
* 旁邊或下方加模型特色說明文字（用簡單框框標明：自動化、無需特徵工程、多尺度設計等）

🟪 頁 5：成果與準確率展示

* 左側放混淆矩陣、準確率條狀圖
* 右側放 t-SNE 結果或其他可視化特徵分佈
* 下方用一句話強調：「>90% 準確率，顯著優於傳統方法」

🟫 頁 6：創新性與應用場景

* 可用「雷達圖」或「表格」強調創新特點（如比較你們 vs 傳統工具）
* 放一張未來應用示意圖（如API預測服務、研究者操作介面）
* 如果能加一些藝術化圖標或插畫點綴會更吸睛

🟨 頁 7：預期效益

* 分兩欄：社會效益（加速研究、精準醫療） & 商業價值（模組化應用、潛在客戶）
* 可以用簡單 icon + 一句話強調每一項效益

🟦 頁 8：附錄與聯絡方式

* 若有GitHub、網站或Demo QR Code，可以放這裡
* 若有Logo或品牌名稱也可設計簡單版型放此頁
* 加上一句口號也會有加分效果（例如：「AI識序，創未來」）