114-1 Machin Learning Final Project

Ming Hsun Wu

1. AI 未來 20 年能力：心理師的「個人化情緒陪伴分身」

我認為在未來 20 年內，AI 有可能實現一項目前技術仍無法做到，但對社會具有重大意義的能力——讓每位心理師都能擁有自己的 AI 分身（Personalized Therapist Avatar）。這個 AI 分身能夠長期陪伴與支持個案，在諮商室外的時間提供個人化、具情緒敏感度的心理輔助。

我的動機是因為友人目前是實習心理師，所以對於「諮商」這方面有較深的認識。**站在個案的角度來看：**這件事情對於現在人有許多困難或是刻板印象，不管是社會與論還是諮商的門檻，對於需要的人都是較難獲得的幫助。而**站在諮商心理師的角度來看：**對於有限的時間和資源，能夠幫助到的個案也是相當有限。再加上「航海王」這部動漫中有個橋段是將自己的意志分割成好幾個不同的個體，分別來完成各自的任務，因此有了這個想法。

為什麼重要？

現今社會對心理諮商的需求逐年提升，但存在多重限制：

- 費用高昂、頻率有限：多數人難以負擔固定諮商。
- 心理師人力不足：一位心理師能服務的個案數有限。
- 社會污名與就醫障礙：許多人即使需要，也不敢走進諮商室。

未來的「心理師 AI 分身」將能做到：

1. 長期情緒建模：持續追蹤個案的情緒模式、壓力來源。
2. 提供治療架構下的介入：如 CBT、ACT、MI 等。
3. 充當心理師與個案的橋樑：在諮商之間協助練習與反思。
4. 偵測高風險訊號並回報心理師：建立安全機制。

這樣的 AI 將使心理照護不再侷限於每週一次的面談，而能延伸到日常生活中，減少心理師的負擔，也讓更多人獲得即時支持。

2. 所需的成分與資源

要實現上述 AI 能力，未來需要整合多種資料、工具、運算環境與學習架構。

2.1 資料

- 長期對話資料：追蹤個案的情緒變化（文字、語音、行為紀錄）。
- 多模態資料：睡眠、生理訊號、語音語調、行為習慣。
- 專業心理治療語料：包含治療師–個案的真實對話、介入方式、標註策略。
- 高風險事件標記：自傷語言、情緒暴衝、極端負面思考。

2.2 工具

- 語言模型（BERT、Transformer 等）：語意理解與情緒辨識。
- Representation Learning：學習個案的個人化情緒向量空間。
- Reinforcement Learning：學習「什麼樣的介入最有效」。
- 心理量表與統計模型：PHQ-9、GAD-7 等做客觀標註。
- 安全機制（Risk Detection Models）：偵測危險語言與突變訊號。

2.3 硬體與環境

- 行動裝置感測器：睡眠、行為、生理訊號。
- 手機 APP 監測：根據使用者使用 3C 產品情況、背景追蹤使用者狀況。
- 邊緣運算或雲端 GPU：支援長期語言模型更新。

2.4 學習架構

- Supervised Learning：情緒分類、意圖判讀。
- Self-Supervised Learning：從大量未標記對話中學語意。
- Unsupervised Learning：找出個案專屬的情緒模式。
- Meta-Learning：快速適應不同個案的風格。

3. 涉及的機器學習類型

心理師 AI 分身屬於一個複合式 AI 系統，需結合三種學習方式：

監督式學習（Supervised Learning）

- 用途：辨識情緒、分析語意、偵測風險。
- 資料來源：已標註語料、心理量表、心理師註解。
- 目標訊號：情緒分類、風險等級、意圖分類。

非監督式學習（Unsupervised Learning）

- 用途：建構每位個案的「個人化情緒模型」。
- 資料來源：個案的長期對話與行為歷史。
- 目標：找出隱含的情緒週期、壓力來源、思考模式。

強化學習（Reinforcement Learning）

- 用途：學習「什麼樣的回應最有幫助」。
- 回饋訊號：個案的情緒改善、語氣緩和、使用者評分、長期 engagement。
- 互動環境：AI 與個案的日常對話。

(ps:參考 GPT)

4. 第一步可實作模型（Toy Model）

為了實現我在 Assinment10 中描述的「心理師 AI 分身」，其核心能力之一是：**根據個案的文字表達，辨識其情緒狀態，並產生具有心理支持性的回應。**這是一個介於語意理解、情緒建模與回饋生成之間的複合任務。我打算先從辨識對話中的情緒開始，我設計了一個較小的「玩具版本」，他能夠判別出對話中的情緒並進行分類當作第一步。

4.1 問題設計

最終目標對應：未來的 AI 分身必須理解個案的情緒，而分類模型正是基礎能力的簡化版本，我將這個 **toy model 設計為判別使用者當下的訊息來辨識使用者的情緒。**

- Input：一段文字（例如使用者當下的訊息）。
- Output：六類情緒之一（anger, fear, joy, love, sadness, surprise）。

4.2 資料

- 各式短句
- 六種類別的情緒標註（anger、fear、joy、love、sadness、surprise）

text	label
i didnt feel humiliated	0
i can go from feeling so hopeless to so damned hopeful just fro	0
im grabbing a minute to post i feel greedy wrong	3
i am ever feeling nostalgic about the fireplace i will know that it	2
i am feeling grouchy	3
ive been feeling a little burdened lately wasnt sure why that was	0
ive been taking or milligrams or times recommended amount an	5
i feel as confused about life as a teenager or as jaded as a year c	4
i have been with petronas for years i feel that petronas has perfc	1
i feel romantic too	2

4.3 模型架構

- **Target function**：$y_i = f(BERT(x_i))$，其中 f 為想找到的目標函數，$BERT$ 為預訓練好的模型進行採樣，x_i 為使用者一句話或一段對話文字，$y_i \in$ {0:anger, 1:fear, 2:joy, 3:love, 4:sadness, 5:surprise}。
- **Hypothesis function**：
  以預訓練的 BERT-base 作為 encoder，後接一個輕量的 classifier（全連接層）
- **Activation function**：Softmax
- **Optimizer**：AdamW
- **Loss Function**：Cross Entropy

4.4 結果

因電腦設備問題，因此只決定跑 50 個 epochs，且資料量縮小為訓練資料只有 4000 筆，若能將 epochs 提升、擴增分類層的複雜程度或提升訓練資料量，分類效果應該會更好。

Epoch 1/50　Train Loss: 1.7011　Validation Loss: 1.6559　Validation Accuracy: 0.3600

Epoch 2/50　Train Loss: 1.6686　Validation Loss: 1.6419　Validation Accuracy: 0.3787

…

Epoch 50/50　Train Loss: 1.5813　Validation Loss: 1.5402　Validation Accuracy: 0.5050