

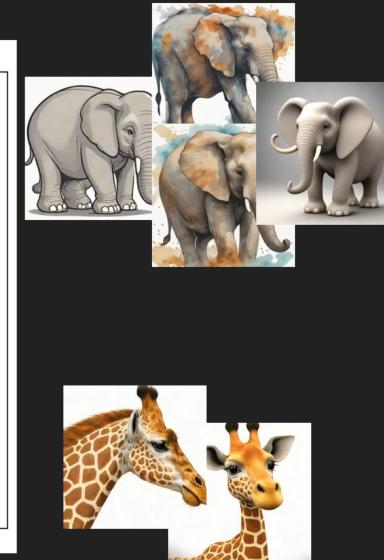
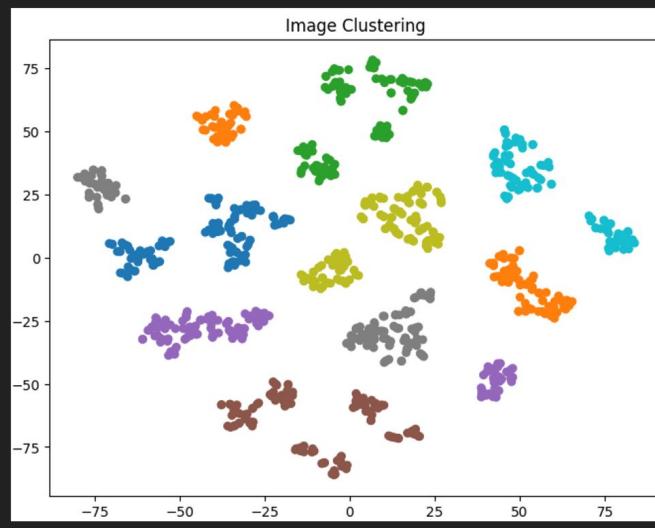
# HW1

# Unsupervised Learning

Introduction to Artificial Intelligence

# Introduction - Unsupervised learning for image classification

- 使用python實作clustering algorithm將影像資料分類
- 請不要使用內建clustering的相關套件 (scikit-learn)
  - 可以使用的套件: numpy跟python預設套件



# Data

Files:

1. features.npy: 5000筆影像資料的embedding (使用CLIP encode)
2. label\_test.npy: 前500筆資料的真實label, 可以評估方法的正確性
3. sample.py: 範例的code, 改成你的clustering algorithm
4. eval.py: 用來評估clustering的好壞
  - 用法: python eval.py {your predicted file path}
5. data: 5000張原始圖片
6. visualize.ipynb: 視覺化結果

\* 原始資料總共有10種類別

# Evaluation metrics

1. Silhouette Coefficient: 衡量 clustering 的品質

- baseline: 0.2

2. Adjusted Rand Index score (ARI): 跟 True label 的匹配率

- Rand index: Measures the similarity between the predicted and true label assignments

$$RI = (TP + FP) / (TP + FP + TN + FN)$$

- Adjusted Rand Index score: Adjusts for the chance grouping and is more reliable than the raw Rand Index.

$$ARI = (RI - \text{Expected\_RI}) / (\max(RI) - \text{Expected\_RI})$$

- reference:

[https://scikit-learn.org/stable/modules/generated/sklearn.metrics.adjusted\\_rand\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.adjusted_rand_score.html)

- baseline: 0.75

# Grading Policy

- Silhouette Coefficient (50%)
  - > baseline: 100 points
- ARI score (50%)

ARI score	Points
Top 10%	100
Top 25%	95
Top 50%	90
Top 75%	85
$s \geq \text{baseline}$	80
$s < \text{baseline}$	0

# Submissions (Deadline: 2025/09/22 23:59 p.m.)

Submit your zipped source code `hw1_{student_id}.zip` to E3. After unzip the file, it should appear a folder `hw1_{student_id}`. The folder should contain the following items:

1. Your code named `{student_ID}.py`
    - The result should be reproduce by running `$python [student_ID].py`
  2. Predicted 5000 label result in numpy format (`{student_ID}.npy`)
- 
- 不用包含其他資料檔, 請預設`features.npy`跟你的執行的預測結果在同一個資料夾底下
  - 繳交範例:  
`$unzip hw1_111222333.zip` (unzip完之後出現以下檔案)  

```
hw1_111222333/
└── 111222333.py
└── 111222333.npy
```

# Rules

1. 任何檔名錯誤影響到助教執行算分一個會扣總分5分。
  2. 禁止使用額外的套件, 若使用助教會依嚴重性斟酌扣分。
  3. 每次clustering結果可能會有小部分差異, 助教demo主要檢查程式是否可以正確執行以及確定可以得出相似的結果即可。
  4. 禁止抄襲, 被發現以0分計算。
- 
- 建議使用Python 3.6以上的版本 (demo環境: python=3.11.7)
  - 如果有任何問題請寄信給助教 曾薈 [chiang.ee11@nycu.edu.tw](mailto:chiang.ee11@nycu.edu.tw)
  - TA hour: 9/17(Wed.) 13:00~15:00 at ED716