List of tasks (Official/Unofficial)

1. Train LSTM with custom and glove embs on a Toxicity detection dataset (binary) **~2 points**
2. Train **2** BERT-alike model (e.g. BERT, RoBERTa) on STS-b dataset (semantic similarity) **~2 points**
3. Create custom RAG mechanism **~3 points**
4. Collect via prompt engineering a list of answers on questions for deep learning providing two answers one gentle and one toxic (Gordon Ramsay style) **~1 point**
5. Tune an LLM using DPO to always act like Gordon Ramsay! **~2 points**

1. Toxicity dataset

https://github.com/Sreyan88/Toxicity-Detection-in-Spoken-Utterances/blob/main/data

Contains already a train.csv, valid.csv and test.csv (ignore trigger_test.csv)

Use all examples (don’t remove for balance)

Compute F1-score and Confusion Matrix for both LSTM with custom and glove embs

Use Tensorflow/Keras

2. STS-b Dataset

https://huggingface.co/datasets/mteb/stsbenchmark-sts

Contains already a train, valid and test set

- dataset = load_dataset("glue", "stsb")

Use all examples

Regression task, labels for STS-B are already a float (0-5 scale)

Compute Pearson and Spearman correlation and Confusion Matrix for **both selected models**

- from scipy.stats import pearsonr, spearmanr

Use Transformers from HF (PyTorch)

4. Preference Alignment Dataset

Gordon Ramsay AIDL tutor

Create a dataset containing **100 examples** each one. It should look like this:

Student Question: Why do we use ReLU instead of sigmoid?

Polite Answer: ReLU avoids saturation in the positive region and reduces vanishing gradients, enabling faster and more stable training.

Ramsay-Style: Stop keep slapping sigmoid everywhere like you're decorating a cake with mayonnaise. Of course the gradients vanish—sigmoid flattens everything! ReLU is there to save the day, because you sure aren't going to do it!

4. Preference Alignment Dataset

Gordon Ramsay AIDL tutor

Add the results to a CSV with **4 columns**:

AIDL_ID, Question, Polite, Ramsay

Check the example here:
https://eclass.uniwa.gr/modules/document/file.php/MSCAIDL112/ramsay_example.csv

Upload the dataset to eclass.

DEADLINE 08/01!!!

Tips: play around with the prompt, be creative, use several interfaces (chatgpt, gemini, claude), **keep the answers short!**