

# Large Language Models Assignment

## Objective

Explore different ways to classify academic abstracts with decoder-only and encoder-only large language models when only a few labeled examples are available for training and testing.

## Data

Training set, validation set, and test set all contains a field *Abstract* that contains the abstracts to classify. The class for each abstract is included in the field *Domain*. The data files also contain the fields *area* and *keywords* which may be used for in-context learning/prompt engineering. The training dataset contains 50 examples, the validation set contains 50 samples, and the test set contains 20 samples.

## Main Task

For each abstract  $a \in test.\text{Abstract}$ , report classification accuracy of different methods defined as:

$$\text{accuracy} = \frac{1}{|test|} \sum_{a \in test.\text{Abstract}} I(\text{model}(a.\text{Domain}) == a.\text{Domain}),$$

where  $|test|$  is the size of the test set,  $I$  is the indicator function and  $I(\text{model}(a.\text{Domain}) == a.\text{Domain})$  means that the model-predicted domain  $\text{model}(a.\text{Domain})$  is equal to the true domain  $a.\text{Domain}$ .

## Deliverables

- This document completed (see highlighted fields)
- Screenshots of the interaction with gpt3.5 for Questions #2 and #3
- The datasets formatted for fine-tuning with Open AI Playground
- The Colab Notebook template (or an equivalent) completed. Requirements for the submitted notebook are that it's easy to read, it runs without errors and it only uses free Colab resources (e.g., cpu or T4). [Link to the notebook](#)

## Contact

For any question about the assignment, please contact me at [layla.elasri@borealisai.com](mailto:layla.elasri@borealisai.com). I can hop on video calls if needed.

## Part I – Using a pre-trained decoder-only model

### Question #1

Using **chatgpt3.5** available at <https://chat.openai.com/> (requires an account but access to the model is free), report the classification accuracy in a **zero-shot setting and without any specific**

**prompt engineering:** the first prompt should only include the task and the subsequent prompts should include the abstracts from the test set.

<h2>Prompt used:</h2>	<p>Given the abstract determine the academic domain it belongs to.</p>	<p>You</p>	<p>Given the abstract determine the academic domain it belongs to.</p>	<p>Possible Domains:</p> <ul style="list-style-type: none"><li>-CS</li><li>-ECE</li><li>-Civil</li><li>-MAE</li><li>-Medical</li><li>-Psychology</li><li>-biochemistry</li></ul>	<p>ChatGPT</p>	<p>To determine the academic domain of an abstract, we need to analyze the content and identify the subject matter. Since you haven't provided the abstract, I can't directly determine the academic domain it belongs to. Could you please provide the abstract so I can assist you further?</p>	<p>Q1 Q2 Q3</p>	<p>Q4 Q5 Q6</p>	<p>Q7 Q8 Q9</p>	<p>Q10 Q11 Q12</p>	<p>Q13 Q14 Q15</p>	<p>Q16 Q17 Q18</p>	<p>Q19 Q20 Q21</p>
-----------------------	--	------------	--	--	----------------	---	-----------------	-----------------	-----------------	--------------------	--------------------	--------------------	--------------------

## Prompt used:

Classification accuracy: The classification accuracy in this case was 0.55 (55%).

(Optional) Observations: A significant challenge I experienced was with the domains of Psychology, and Civil, because some of the abstracts for these two domains in particular produced outputs such as Education, and Environment respectively. Getting from these domains into one of the 7 domains for classification was challenging, and required various prompts.

## Question #2

Continuing with **chatgpt3.5**, experiment with prompt engineering **on the training set** and try to improve **test accuracy**. By this, I mean that you can give training examples, explain reasoning on the training set, etc. and then prompt chatGPT again on the test set. Samples from the test set are only to be used as final cues, i.e., they cannot be used as in-context learning examples or for chain-of-thoughts reasoning, nor any other type of prompt engineering. The goal here is to use training examples to teach chatGPT the task and then evaluate it again on the test set to test its generalization abilities.

Prompts used: please attach screenshots of the entire interaction

### You

I will present you with several abstracts along with their respective domains. Your task is to learn from these and then the new abstracts provided classify them into one the domains.

The domains are:  
-CS  
-Civil  
-Medical  
-Psychology  
-MAE  
-ECE  
-biochemistry

#### Example 1:

**Abstract:** Aims: This study aimed to develop and test the feasibility, acceptability and initial effectiveness of a three-session psychosocial group intervention to reduce hepatitis C risk behaviours and increase hepatitis C transmission knowledge among women who inject drugs in five European cities/towns. Methods: An uncontrolled, field effectiveness study of a psychosocial group intervention. Hepatitis C virus (HCV) transmission knowledge, sexual and drug risk behaviours and depressive symptoms were assessed at baseline and one-month post-intervention. Intention-to-treat analyses were conducted. Findings: One-month post-intervention, a significant increase was reported in HCV transmission knowledge and in the number of new and unused needles/syringes used to inject. There were significant reductions in the sharing of spoons/containers for mixing that had been used by someone else, sharing of filters, cookers, spoons or water with someone who was hepatitis C positive and the use of alcohol swabs following injection. Conclusions: The intervention showed promising results in reducing some hepatitis C injecting risk behaviours and increasing hepatitis C transmission knowledge among women who inject drugs. These preliminary findings suggest that it is feasible to deliver the intervention in drug treatment settings, and that the intervention was acceptable to both participants and staff.

Domain: Medical

#### Example 2:

**Abstract:** Mobile agent technology is becoming more popular and has been implemented in many distributed computing domains. Several research have been conducted to address its challenges including two most important ones which are agent spawning and agent mobility. This paper discusses the issues of mobile agent technology, the background of mobile agent cloning and spawning, agent mobility as well as the problems faced by many researchers in their research on mobile agents. The paper finally proposes new agent spawning and mobility models to resolve some of the researchers' problems.

Domain: CS

#### Example 3:

**Abstract:** This paper presents an educational software tool, called wtControlGUI, whose main purpose is to show the applicability and performance of different decoupling control strategies in wind turbines. Nowadays, wind turbines are a very important field in control engineering. Therefore, from an educational point of view, the tool also aims to improve the learning of multivariable control concepts applied on this field. In addition, wtControlGUI allows for testing and controlling of a lab-scale system which emulates the dynamic response of a large-scale wind turbine. The designed graphical user interface essentially allows simulation and experimental testing of decoupling networks and other multivariable methodologies, such as robust or decentralized control strategies. The tool is available for master degree students in control engineering. A survey was performed to evaluate the effectiveness of the proposed tool when it is used in educational related tasks. © 2016 Wiley Periodicals, Inc.

Domain: ECE

#### Example 4:

**Abstract:** Different researchers have reported positive, null, and negative relationships between confidence and accuracy in reports from memory. One possible reason for this paradox is the variety of materials used across experiments, but the two experiments reported in this article show that positive and negative confidence-accuracy correlations can be observed using a single procedure and the same materials. Subjects studied words from semantic categories and then took a recognition test while making confidence ratings. For previously studied items, positive correlations between confidence and accuracy were obtained using three different measures. Yet when confidence-accuracy correlations were assessed for unstudied items from studied categories, the correlations were zero or negative. The critical factors in determining when negative correlations will be found are the similarity of lures to presented items and the type of analysis used. These results indicate that one should be cautious about relying on confidence of recognition when rememberers must decide among highly similar events.

Domain: Psychology

### Example 5:

**Abstract:** In this paper, we focus on tackling the problem of automatic accurate localization of detected objects in high-resolution remote sensing images. The two major problems for object localization in remote sensing images caused by the complex context information such images contain are achieving generalizability of the features used to describe objects and achieving accurate object locations. To address these challenges, we propose a new object localization framework, which can be divided into three processes: region proposal, classification, and accurate object localization process. First, a region proposal method is used to generate candidate regions with the aim of detecting all objects of interest within these images. Then, generic image features from a local image corresponding to each region proposal are extracted by a combination model of 2-D reduction convolutional neural networks (CNNs). Finally, to improve the location accuracy, we propose an unsupervised score-based bounding box regression (USB-BBR) algorithm, combined with a nonmaximum suppression algorithm to optimize the bounding boxes of regions that detected as objects. Experiments show that the dimension-reduction model performs better than the retrained and fine-tuned models and the detection precision of the combined CNN model is much higher than that of any single model. Also our proposed USB-BBR algorithm can more accurately locate objects within an image. Compared with traditional features extraction methods, such as elliptic Fourier transform-based histogram of oriented gradients and local binary pattern histogram Fourier, our proposed localization framework shows robustness when dealing with different complex backgrounds.

Domain: Civil

### Example 6:

**Abstract:** A full-length Beet black scorch virus (BBSV) cDNA clone (pUBF52) was constructed by RT-PCR. The clone contains an upstream T7 RNA polymerase promoter designed for in vitro transcription of infectious RNAs from the linearized plasmid that faithfully represent the viral cDNA. Leaves of *Chenopodium amaranticolor* inoculated with in vitro transcripts developed the same symptoms and disease phenotype as the wild type virus. The presence of BBSV RNA and coat protein in the leaves was confirmed respectively by Northern blotting and Western blotting. Comparisons of specific immunoreactions between the expression product of the BBSV p24 gene in *E. coli* and antiserum against purified BBSV virions demonstrated that the p24 gene encodes the coat protein. Based on the sequence of the pUBF52 cDNA, a frame-shift mutant and two deletion mutants were generated. One of the deletions encompasses the entire CP ORF and the other truncates 174 amino acids from the central region of the protein. Transcripts derived from the frame-shift CP mutant, which terminates the CP after the first 23 amino acids, elicited the same symptom phenotype and levels of RNA accumulation as the wild type virus, but the leaves infected with the CP deletion mutants exhibited greatly reduced RNA accumulation. In addition, leaves inoculated with in vitro transcripts of the mutant in which the entire CP gene was deleted had lower local lesions than wild type virus transcripts. Two expression vectors, pBGFP and pBGUS, were constructed by fusing the GFP and GUS genes to the 23 N-terminal amino acids of the CP gene, respectively. Leaves infected with in vitro transcripts of pBGFP and pBGUS exhibited expression of GFP and GUS proteins as assessed by laser confocal microscopy and histochemical staining, respectively. The high levels of expression of the GFP and GUS proteins provide tools that can be used for studies of replication and movement of the virus, and indicate that BBSV has considerable biotechnology potential as a plant virus expression vehicle.

Domain: biochemistry

### Example 7:

**Abstract:** A nonlinear dynamic coupled model for hydropower station system, which contains the model of water-carriage system, water turbine system, speed governor system, generator's electromagnetic system, grid, shaft system of hydroelectric generating set, as well as the powerhouse, is established in this paper. Firstly, the simultaneous differential equations for coupled hydraulic-mechanical-electric transient process are set up based upon the theories of hydraulics, electrical machinery, etc., while the coupled structural models for shaft system of unit and powerhouse are built by means of finite element method. Secondly, a new method for investigating nonlinear dynamic properties of structures influenced by coupled hydraulic-mechanical-electric factors in different conditions is introduced with the help of user-programmable features from Ansys software. Finally, in order to verify the rationality, several numerical calculation methods are used to study the starting-up process of hydropower station. The results indicate that the model presented in this paper is adoptable for simulating specified condition and reflect the nonlinear dynamic characteristics of hydropower station comprehensively. In addition, the model can also be used to assess the operation safety and predict the structures reliability of hydropower station system, so as to provide some profitable reference for dynamic regulation during limited and transient conditions for hydropower station.

Domain: MAE

### ChatGPT

Based on the abstracts provided, here's the classification of the new abstracts into the respective domains:

#### 1. Abstract:

A new algorithm for image compression is proposed in this paper. The algorithm utilizes a combination of wavelet transformation and vector quantization techniques to achieve high compression ratios while maintaining acceptable image quality.

Domain: CS

#### 2. Abstract:

This study investigates the effects of different construction materials on the structural integrity of bridges under various loading conditions. Experimental tests were conducted on bridge models constructed using concrete, steel, and composite materials, and the results were analyzed to assess the performance and durability of each material.

Domain: Civil

#### 3. Abstract:

An analysis of the impact of social media usage on adolescent mental health is presented in this paper. A survey was conducted among teenagers to examine their social media habits and assess their levels of stress, anxiety, and depression. The findings reveal correlations between excessive social media use and negative mental health outcomes.

Domain: Psychology

### You

Consider a few more examples:

#### Example 1:

Abstract: The dopamine receptor-D4 and the dopamine transporter have been investigated for their role in attention deficit hyperactivity disorder (ADHD) in children. Reports of their genetic association with ADHD have shown mixed results. The aim of the study was to evaluate the association of variable number tandem repeats (VNTRs) of the DRD4 and DAT1 genes with ADHD in children. A pilot 1:1 case control study, with 44 clinically confirmed ADHD cases and 44 age/gender matched healthy controls, was conducted at a tertiary care centre in Mumbai. Variable number tandem repeats of DRD4 exon 3, DAT1 intron 8 and 3'UTR were genotyped by PCR-AGE. Several allele repeats of the genes were observed in the screened subjects. Statistical significance was observed for the I0R/I0R genotype of the DAT1 3'UTR VNTR between cases and controls.

Domain: Psychology

#### Example 2:

Abstract: [Purpose] The aim of this study was to analyze stabilometry in athletes during an indoor season in order to determine whether injured athletes show different stabilometric values before injury than non-injured athletes in two different training periods (volume and pre-competition periods). [Subjects] The subjects were 51 athletes from Unicaja athletic club who trained regularly. [Methods] At the end of the preseason and volume periods, athletes were subjected to bipodal and monopodal stabilometry. In addition, all injuries happening in the periods after performing stabilometry (volume and pre-competition periods) were tracked. [Results] Variance analysis of bipodal stabilometric measurements taken at the end of the preseason period showed that athletes with higher values for the center-of-pressure spread variables suffered injuries during the volume period. The right-leg monopodal stabilometric measurements taken at the end of the volume period showed that athletes with higher values in the center-of-pressure position variables suffered injuries during the pre-competition period. [Conclusion] Athletes showing the worst values for center-of-pressure spread variables are more prone to sports injuries in the subsequent training period. In monopodal measurements, athletes with poorer mediolateral stability were more prone to injuries in the subsequent training period.

Domain: Medical

### Example 3:

Abstract: 1. Angiotensin (Ang) II plays a major role in vascular remodelling. Matrix metalloproteinases (MMPs) and their tissue inhibitors (TIMPs) are involved in the tissue remodelling processes. The aim of the present study was to investigate whether AngII modulates TIMP-2 expression in rat aortic smooth muscle cells in vivo. 2. Angiotensin II (200 ng/kg per min, s.c.) or AngII + losartan (10 mg/kg per day, s.c.) or normal saline was administered continuously by osmotic minipumps to Sprague-Dawley rats for 1 week. In addition, the effect of endogenous AngII on TIMP-2 expression was evaluated in renovascular hypertensive rats (two kidney, one clip (2K1C) and one kidney, one clip (1K1C) models). Control rats (sham 2K1C and sham 1K1C rats) underwent sham-clipping of the left renal artery. At the end of the treatment, plasma renin activity was measured by radioimmunoassay; aortic TIMP-2 mRNA expression was evaluated by real-time polymerase chain reaction and/or northern blotting and protein expression was evaluated by immunohistochemistry. Systolic blood pressure (SBP) was measured twice a week by the tail-cuff method. 3. Exogenous AngII administration produced the expected increase in SBP ( $P = 0.02$ ) compared with the control saline-treated group. The increase in SBP was abolished in AngII + losartan-treated rats. Administration of AngII caused a significant increase in TIMP-2 expression ( $P = 0.01$ ) in rat aortic smooth muscle cells that was abolished in AngII + losartan-treated rats. In renovascular hypertensive rats, SBP was higher ( $P < 0.0001$ ) in 2K1C and 1K1C rats compared with the corresponding sham-operated rats. Plasma renin activity was higher ( $P < 0.01$ ) in 2K1C rats compared with the other groups. The expression of TIMP-2 was significantly ( $P < 0.05$ ) increased only in 2K1C rats. 4. Our *in vivo* data demonstrate that exogenous and endogenous AngII increases TIMP-2 expression in rat aortic smooth muscle cells. This effect is not dependent on the AngII-induced increase in blood pressure and is mediated by angiotensin AT(1) receptors.

Domain: biochemistry

### Example 4:

Abstract: How and at what spatial scale(s) animals change their movements in relation to their environment is central to several topics in ecology and conservation, including foraging ecology, habitat selection and dispersal. A method (first-passage time analysis, FPT) has recently been proposed to measure changes in movements through the landscape, as an index of search effort at the pertinent spatial scales. This method seems largely applicable to an increasing number of studies using satellite, radio-tracking or global positioning system (GPS), but its limits have not yet been assessed. 2. Here I used several movement simulations to examine the ability of FPT analysis to detect area-restricted search (ARS) according to different changes in movements, different patch structures and tracking accuracy. 3. FPT analysis was able to detect changes in movements when both speed and sinuosity changed, or when the animal reacted to patch boundaries. It was also able to detect ARS within the same path at several spatial scales in patches (nested or not) of different sizes. 4. Tracking accuracy affected the detection of ARS by FPT analysis. With the widely used Argos system, a minimum of 13 locations in effective ARS was necessary to detect this behaviour; when seven velocity filtering was applied. Similarly, spatial error in location affected the estimation of the ARS scale value, but the application of velocity filtering reduced this effect. 5. Comparisons between a real GPS track and pathways simulating the Argos error showed that the time-sampling rate of locations (due to satellite-pass frequency) decreased the probability of detecting ARS at small scales (50% across the whole range of scales). A velocity filter enabled significant reductions in this effect. 6. Synthesis and application. Within limits, FPT analysis is highly suitable for animal movement analysis, either to quantify habitat use, or to determine the scale most relevant for describing an ecological system or factors affecting movement decisions. In anticipation of increasing applications of FPT analysis in applied ecology, I provide recommendations for the use of the technique with several tracking methods.

Domain: ECE

### ChatGPT

Based on the abstracts provided, here's the classification of the new abstracts into the respective domains:

#### 1. Abstract:

A new algorithm for image compression is proposed in this paper. The algorithm utilizes a combination of wavelet transformation and vector quantization techniques to achieve high compression ratios while maintaining acceptable image quality.

Domain: CS

#### 2. Abstract:

This study investigates the effects of different construction materials on the structural integrity of bridges under various loading conditions. Experimental tests were conducted on bridge models constructed using concrete, steel, and composite materials, and the results were analyzed to assess the performance and durability of each material.

Domain: Civil

#### 3. Abstract:

An analysis of the impact of social media usage on adolescent mental health is presented in this paper. A survey was conducted among teenagers to examine their social media habits and assess their levels of stress, anxiety, and depression. The findings reveal correlations between excessive social media use and negative mental health outcomes.

Domain: Psychology

You  
Now I will start providing the abstracts and classify them into one of the 7 domains

ChatGPT  
Sure, please provide the abstracts, and I'll classify them into the respective domains.

You  
XML is a pervasive technology for representing and accessing semi-structured data. XPath is the standard language for navigational queries on XML documents and there is a growing demand for its efficient processing. In order to increase the efficiency in executing four navigational XML query primitives, namely descendants, ancestors, children and parent, we introduce a new paradigm where traditional approaches based on the efficient traversing of nodes and edges to reconstruct the requested subtrees are replaced by a brand new one based on basic set operations which allow us to directly return the desired subtree, avoiding to create it passing through nodes and edges. Our solution stems from the NESTED Sets for Object hierArchies (NEASTOR) formal model, which makes use of set-inclusion relations for representing and providing access to hierarchical data. We define in-memory efficient data structures to implement NESTOR, we develop algorithms to perform the descendants, ancestors, children and parent query primitives and we study their computational complexity. We conduct an extensive experimental evaluation by using several datasets: digital archives (EAD collections), INEX 2009 Wikipedia collection, and two widely-used synthetic datasets (XMark and XGen). The solution stems from the NESTOR-based data structures and query primitives consistently outperform state-of-the-art solutions for XPath processing at execution time and they are competitive in terms of both memory occupation and pre-processing time. (C) 2015 Elsevier Ltd. All rights reserved.

ChatGPT  
Domain: CS (Computer Science)

This abstract discusses the efficiency of XPath processing, a technology closely related to representing and accessing semi-structured data, which falls under the realm of computer science. The abstract talks about introducing a new paradigm for processing navigational XML queries more efficiently, which involves developing algorithms and data structures. Additionally, it mentions conducting experiments and evaluating performance using various datasets, which are typical practices in computer science research. Therefore, this abstract fits well within the domain of computer science.

You  
The integration of renewable energy sources increases the complexity in maintaining the power grid. In particular, the highly dynamic nature of generation and consumption demands for a better utilization of energy resources, which seen the cost of storage infrastructure, can only be achieved through demand-response. Accordingly, the availability of energy and potential overload situations can be reflected using a price signal. The effectiveness of this mechanism arises from the flexibility of device operation, which is nevertheless heavily reliant on the exchange of information between the grid and its consumers. In this paper, we investigate the capability of an interactive energy management system to timely inform users on energy usage, in order to promote an optimal use of local resources. In particular, we analyze data being collected in several households in Italy and Austria to gain insights into usage behavior and drive the design of more effective systems. The outcome is the formulation of energy efficiency policies for residential buildings, as well as the design of an energy management system, consisting of hardware measurement units and a management software. The Mjolnir framework, which we release for open use, provides a platform where various feedback concepts can be implemented and assessed. This includes widgets displaying disaggregated and aggregated consumption information, as well as daily production and tailored advices. The formulated policies were implemented as an advisor widget able to autonomously analyze usage and provide tailored energy feedback. The advisor is estimated leading to a potential of 34% of savings using measurement data from the GREEND dataset.

ChatGPT  
Domain: ECE (Electrical and Computer Engineering)

This abstract focuses on energy management systems and their role in optimizing energy usage within the context of integrating renewable energy sources into the power grid. It discusses the technical aspects of designing and implementing such systems, including hardware measurement units and management software. Additionally, it mentions analyzing data collected from households and formulating energy efficiency policies, which are topics typically addressed within the realm of electrical and computer engineering. Therefore, this abstract fits within the domain of electrical and computer engineering.

You  
The original free-swinging piston engine with the oscillating rotating generator distinguishes by its compactness, minimum friction losses, full balanced mechanical system, possibility of four-stroke operation of the engine, possibility of the multifunctional operation of the engine, optimally magnetic circuit of the electrical generator, etc. Because the engine can optimally operate over whole range of the output power, it enables to improve electrical system of the hybrid vehicle by usage of supercapacitors; and optimised energy flows. In this way the competitive hybrid electric vehicle can be created.

ChatGPT  
Domain: MAE (Mechanical and Aerospace Engineering)

This abstract discusses the design and features of a free-swinging piston engine with an oscillating rotating generator, focusing on its mechanical aspects such as compactness, minimum frictional losses, balanced mechanical system, and its ability to operate as a four-stroke engine and use multiple fuels. These topics fall within the domain of mechanical engineering. Additionally, it mentions optimizing energy flows and improving the electrical system of hybrid vehicles, which involves the integration of mechanical and electrical systems, aligning with the interdisciplinary nature of mechanical and aerospace engineering. Therefore, this abstract fits within the domain of mechanical and aerospace engineering.

You  
Barriers to access and long-term complications remain a challenge in transplantation. Further advancements may be achieved through research priority setting with patient engagement to strengthen its relevance. We evaluated research priority setting in solid organ transplantation and described stakeholder priorities using a Delphi technique in October 2016. We synthesized the findings across 12 studies (n = 177 participants) and identified 32 research priorities (liver 17 [7 (25%)], liver (3 (1%), lung (1 (4%), pancreas (1 (4%), and non-specific organ transplantation (7 (25%)) using consensus conferences, expert panel meetings, workshops, surveys, focus groups, interviews, and the Delphi technique. Nine (32%) reported patient involvement. The 339 research priorities addressed the following: organ donation (43 priorities (14 studies)); waitlisting and allocation (43 (10 studies); heterocompatibility and immunology (31 (8 studies); immunosuppression (10 (3 studies); donor selection (10 (3 studies); recipient (non-graft-related) complications (8 (4 studies); reproduction (1 (1 study); psychosocial and lifestyle (49 (7 studies); and disparities in access and outcomes (10 (4 studies)). The priorities identified were broad but only one-third of initiatives engaged patients/caregivers, and details of the process were lacking. Setting research priorities in an explicit manner with patient involvement can guide investment toward the shared priorities of patients and health professionals.

ChatGPT  
Domain: Medical

This abstract discusses research priority setting in the field of solid organ transplantation, focusing on identifying barriers to access and long-term complications. It mentions evaluating stakeholder priorities, including patients, and synthesizing findings from various studies. Additionally, it addresses the importance of patient engagement in research priority setting to ensure research and guide investment towards shared priorities. These topics are pertinent to the medical field, particularly in the context of transplantation and improving patient outcomes. Therefore, this abstract fits within the domain of medicine.

You  
This paper is to present a technological solution for implementing a force feedback haptic interface in the context of virtual reality interaction applications. The real-time haptic algorithms are implemented in TwinCAT runtime environment to control drivers via EtherCAT fieldbus. A 3-dimensional visualization model for the virtual reality interaction is developed through Virtual Reality Toolbox in Matlab/Simulink and interconnected with TwinCAT software through the Automation Device Specification communication protocol. The using of Matlab/Simulink with model-based programming method instead of the C++ programming language allows researchers in haptic domain to focus more on the control engineering issues than programming skills. Some experimental tests and verification of the haptic algorithms with visualization model are presented.

ChatGPT  
Domain: ECE (Electrical and Computer Engineering)

This abstract describes the development of a technological solution for implementing a force feedback haptic interface in virtual reality interaction applications. It involves real-time haptic algorithms implemented in TwinCAT runtime environment, controlling drivers via EtherCAT fieldbus. Additionally, it mentions developing a 3-dimensional visualization model using Virtual Reality Toolbox in Matlab/Simulink and interconnecting it with TwinCAT software. These topics

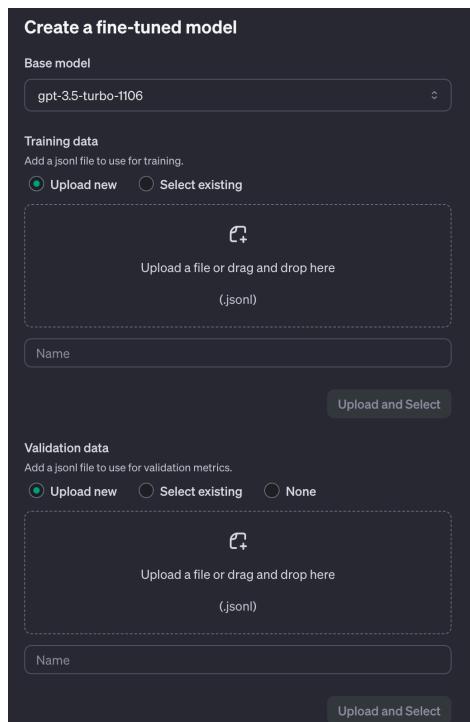


Classification accuracy: In this case with some level of training examples the classification accuracy attained was again 0.55 (55%).

(Optional) Observations: In this case the number of times gpt 3.5 was classifying abstracts into one of the 7 domains was higher, so there was little prompting I had to do to ensure the model was able to classify abstracts correctly, but again one abstract of Psychology I got into the domain of Education as output which was tedious to mitigate.

### Question #3

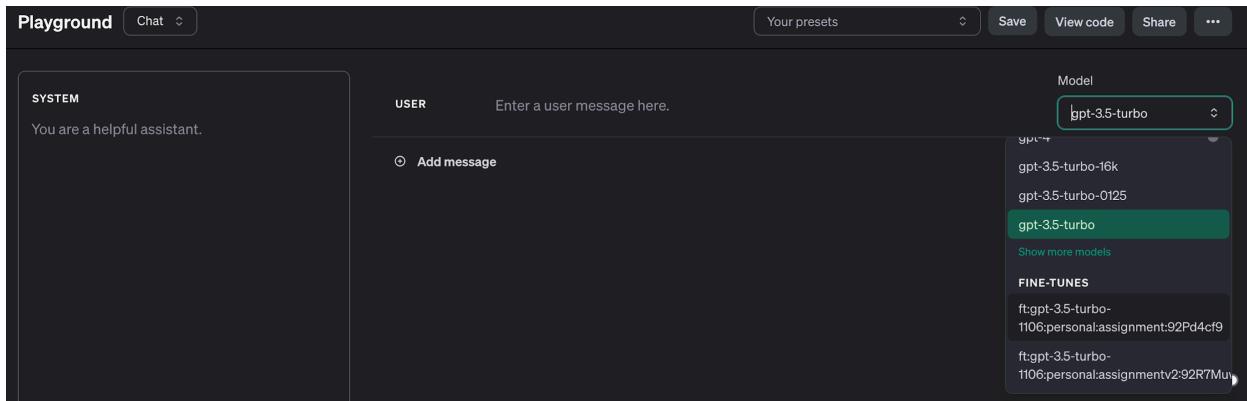
Using the fine-tuning functionality from [OpenAI Playground](#), use the training set to fine-tune **gpt3.5-turbo-1106** on the training set. Use the validation set for this fine-tuning, the interface looks like this:



Before uploading the training and validation datasets, you will need to format them for the fine-tuning task. Please **follow the formatting instructions for gpt3.5-turbo described [here](#)**. Fill in the provided [Google colab template](#) to write the code for the data formatting (see Section Question #3: *Prepare the data for fine-tuning using OpenAI Playground* in the notebook).

Note that this is a **paying service**. Each fine-tuning job **costs less than one dollar** (about \$0.75) and only one job is needed but if you would like to experiment with different parameters, a **budget of \$5** is available for this task. If you have any issues or need to go above budget, please reach out to me beforehand so I can help and ensure that you are able to perform the task.

Once the fine-tuning is done, you can access the fine-tuned model through the [playground](#) under Model/fine-tunes:



**Repeat Questions 1 and 2 with this fine-tuned model.**

Without Prompt-engineering

A screenshot of the playground interface showing a classification task. The 'USER' box contains the instruction: 'Given the following abstracts classify them into one of the following 7 domains:' followed by a list of possible domains: - Civil - CSE - Medical - Psychology - Biochemistry - MAE. The 'ASSISTANT' box contains the following responses:

- 1. The Effect of Music Therapy on Anxiety Levels in Hospitalized Patients  
Domain: Medical
- 2. Neural Network Models for Sentiment Analysis in Social Media Texts  
Domain: CS
- 3. The Impact of Air Pollution on Urban Infrastructure Deterioration  
Domain: Civil
- 4. Development of Wearable Sensors for Monitoring Cardiac Activity  
Domain: ECE
- 5. The Role of Genetics in the Development of Schizophrenia  
Domain: Psychology
- 6. Investigating the Molecular Pathways of Cancer Metastasis  
Domain: Biochemistry
- 7. Design and Optimization of Unmanned Aerial Vehicles for Surveillance Applications  
Domain: MAE

Prompt used:



**Classification accuracy:** In this case the classification accuracy produced is 0.55 (55%).

(Optional) Observations: In this case the fine tuned gpt 3.5 model was certainly able to perform better than the gpt 3.5 model without fine tuning; however, still with the domain of Psychology there were issues as with an abstract it was classifying it as Education, and causing difficulty to associate Education with one of the 7 domains.

## With prompt-engineering

### Prompts used: please attach screenshots of the entire interaction

**USER**

I will present you with several abstracts along with their respective domains. Your task is to learn from these and then the new abstracts provided classify them into one the domains.

The domains are:  
-CS  
-Civil  
-Medical  
-Psychology  
-MAE  
-ECE  
-biochemistry

Example 1:  
Abstract: This study aimed to develop and test the feasibility, acceptability and initial effectiveness of a three-session psychosocial group intervention to reduce hepatitis C risk behaviours and increase hepatitis C transmission knowledge among women who inject drugs in five European cities/towns. Methods: An uncontrolled, field effectiveness study of a psychosocial group intervention. Hepatitis C virus (HCV) transmission knowledge, sexual and drug risk behaviours and depressive symptoms were assessed at baseline and one-month post-intervention. Intention-to-treat analyses were conducted. Findings: One-month post-intervention, a significant increase was reported in HCV transmission knowledge and in the number of new and unused needles/syringes used to inject. There were significant reductions in the sharing of spoons/containers for mixing that had been used by someone else, sharing of filters, cookers, spoons or water with someone who was hepatitis C positive and the use of alcohol swabs following injection. Conclusions: The intervention showed promising results in reducing some hepatitis C injecting risk behaviours and increasing hepatitis C transmission knowledge among women who inject drugs. These preliminary findings suggest that it is feasible to deliver the intervention in drug treatment settings, and that the intervention was acceptable to both participants and staff.  
Domain: Medical

Example 2:  
Abstract: Mobile agent technology is becoming more popular and has been implemented in many distributed computing domains. Several research have been conducted to address its challenges including two most important ones which are agent spawning and agent mobility. This paper discusses the issues of mobile agent technology, the background of mobile agent cloning and spawning, agent mobility as well as the problems faced by many researchers in their research on mobile agents. The paper finally proposes new agent spawning and mobility models to resolve some of the researchers' problems.  
Domain: CS

Example 3:  
Abstract: This paper presents an educational software tool, called wtControlGUI, whose main purpose is to show the applicability and performance of different decoupling control strategies in wind turbines. Nowadays, wind turbines are a very important field in control engineering. Therefore, from an educational point of view, the tool also aims to improve the learning of multivariable control concepts applied on this field. In addition, wtControlGUI allows for testing and controlling of a lab-scale system which emulates the dynamic response of a large-scale wind turbine. The designed graphical user interface essentially allows simulation and experimental testing of decoupling networks and other multivariable methodologies, such as robust or decentralized control strategies. The tool is available for master degree students in control engineering. A survey was performed to evaluate the effectiveness of the proposed tool when it is used in educational related tasks. (C) 2016 Wiley Periodicals, Inc.  
Domain: ECE

Example 4:  
Abstract: Different researchers have reported positive, null, and negative relationships between confidence and accuracy in reports from memory. One possible reason for this paradox is the variety of materials used across experiments, but the two experiments reported in this article show that positive and negative confidence-accuracy correlations can be observed using a single procedure and the same materials. Subjects studied words from semantic categories and then took a recognition test while making confidence ratings. For previously studied items, positive correlations between confidence and accuracy were obtained using three different measures. Yet when confidence-accuracy correlations were assessed for unstudied items from studied categories, the correlations were zero or negative. The critical factors in determining when negative correlations will be found are the similarity of lures to presented items and the type of analysis used. These results indicate that one should be cautious about relying on confidence of recognition when rememberers must decide among highly similar events.  
Domain: Psychology

Example 5:  
Abstract: In this paper, we focus on tackling the problem of automatic accurate localization of detected objects in high-resolution remote sensing images. The two major problems for object localization in remote sensing images caused by the complex context information such images contain are achieving generalizability of the features used to describe objects and achieving accurate object locations. To address these challenges, we propose a new object localization framework, which can be divided into three processes: region proposal, classification, and accurate object localization process. First, a region proposal method is used to generate candidate regions with the aim of detecting all objects of interest within these images. Then, generic image features from a local image corresponding to each region proposal are extracted by a combination model of 2-D reduction convolutional neural networks (CNNs). Finally, to improve the location accuracy, we propose an unsupervised score-based bounding box regression (USB-BBR) algorithm, combined with a nonmaximum suppression algorithm to optimize the bounding boxes of regions that detected as objects. Experiments show that the dimension-reduction model performs better than the retrained and fine-tuned models and the detection precision of the combined CNN model is much higher than that of any single model. Also our proposed USB-BBR algorithm can more accurately locate objects within an image. Compared with traditional features extraction methods, such as elliptic Fourier transform-based histogram of oriented gradients and local binary pattern histogram Fourier, our proposed localization framework shows robustness when dealing with different complex backgrounds.  
Domain: Civil

Example 6:  
Abstract: A full-length Beet black scorch virus (BBSV) cDNA clone (pUBFS2) was constructed by RT-PCR. The clone contains an upstream T7 RNA polymerase promoter designed for in vitro transcription of infectious RNAs from the linearized plasmid that faithfully represent the viral cDNA. Leaves of Chenopodium amaranthicola inoculated with in vitro transcripts developed the same symptoms and disease phenotype as the wild type virus. The presence of BBSV RNA and coat protein in the leaves was confirmed respectively by Northern blotting and Western blotting. Comparisons of specific immunoreactions between the expression product of the BBSV p24 gene in *E. coli* and antiserum against purified BBSV virions demonstrated that the p24 gene encodes the coat protein. Based on the sequence of the pUBFS2 cDNA, a frame-shift mutant and two deletion mutants were generated. One of the deletions encompasses the entire CP ORF and the other truncates 174 amino acids from the central region of the protein. Transcripts derived from the frame-shift CP mutant, which terminates the CP after the first 23 amino acids, elicited the same symptom phenotype and levels of RNA accumulation as the wild type virus, but the leaves derived from the CP deletion mutant exhibited greatly reduced RNA accumulation. In addition, leaves inoculated with in vitro transcripts of the mutant in which the entire CP gene was deleted had lower local lesions than wild type virus transcripts. Two expression vectors, pBGFP and pGUS, were constructed by fusing the GFP and GUS genes to the 23 N-terminal amino acids of the CP gene, respectively. Leaves infected with in vitro transcripts of pBGFP and pGUS exhibited expression of GFP and GUS proteins as assessed by laser confocal microscopy and histochemical staining, respectively. The high levels of expression of the GFP and GUS proteins provide tools that can be used for studies of replication and movement of the virus, and indicate that BBSV has considerable biotechnology potential as a plant virus expression vehicle.  
Domain: biochemistry

Example 7:  
Abstract: A nonlinear dynamic coupled model for hydropower station system, which contains the model of water-carriage system, water turbine system, speed governor system, generator's electromagnetic system, grid, shaft system of hydroelectric generating set, as well as the powerhouse, is established in this paper. Firstly, the simultaneous differential equations for coupled hydraulic-mechanical-electric transient process are set up based upon the theories of hydraulics, electrical machinery, etc., while the coupled structural models for shaft system of unit and powerhouse are built by means of finite element method. Secondly, a new method for investigating nonlinear dynamic properties of structures influenced by coupled hydraulic-mechanical-electric factors in different conditions is introduced with the help of user-programmable features from Ansys software. Finally, in order to verify the rationality, several numerical calculation methods are used to study the starting-up process of hydropower station. The results indicate that the model presented in this paper is adoptable for simulating specified condition and reflect the nonlinear dynamic characteristics of hydropower station comprehensively. In addition, the model can also be used to assess the operation safety and predict the structures reliability of hydropower station system, so as to provide some profitable reference for dynamic regulation during limited and transient conditions for hydropower station.  
Domain: MAE

**ASSISTANT**

Abstract 1:  
A study was conducted to examine the impact of a psychosocial group intervention on reducing hepatitis C risk behaviors and increasing knowledge of hepatitis C transmission among women who inject drugs. The intervention showed positive results in improving knowledge and reducing risky behaviors related to hepatitis C transmission.  
Domain: Medical

Abstract 2:  
This paper discusses the challenges and solutions related to mobile agent technology, focusing on agent spawning and mobility. New models for agent spawning and mobility are proposed to address the issues faced by researchers in the field.  
Domain: CS

Consider a few more examples:

**Example 1:**  
Abstract: The dopamine receptor-D4 and the dopamine transporter have been investigated for their role in attention deficit hyperactivity disorder (ADHD) in children. Reports of their genetic association with ADHD have shown mixed results. The aim of the study was to evaluate the association of variable number tandem repeats (VNTRs) of the DRD4 and DAT1 genes with ADHD in children. A pilot T1 case control study, with 44 clinically confirmed ADHD cases and 44 age/gender matched healthy controls, was conducted at a tertiary care centre in Madrid. Variable number tandem repeats of DRD4 exon 3, DAT1 intron 8 and 5'UTR were genotyped by PCR-AGE. Several allele repeats of the genes were observed in the screened subjects. Statistical significance was observed for the 10R/10R genotype of the DAT1 3'UTR VNTR between cases and controls.  
Domain: Psychology

**Example 2:**  
Abstract: [Purpose] The aim of this study was to analyze stabilometry in athletes during an indoor season in order to determine whether injured athletes show different stabilometric values before injury than non-injured athletes in two different training periods (volume and pre-competition periods). [Subjects] The subjects were 51 athletes from Unicaja athletic club who trained regularly. [Methods] At the end of the preseason and volume periods, athletes were subjected to bipodal and monopodal stabilometry. In addition, all injuries happening in the periods after performing stabilometry (volume and pre-competition periods) were tracked. [Results] Variance analysis of bipodal stabilometric measurements taken at the end of the preseason period showed that athletes with higher values for the center-of-pressure spread variables suffered injuries during the volume period. The right-leg monopodal stabilometric measurements taken at the end of the volume period showed that athletes with higher values in the center-of-pressure position variables suffered injuries during the pre-competition period. [Conclusion] Athletes showing the worst values for center-of-pressure spread variables are more prone to sports injuries in the subsequent training period. In monopodal measurements, athletes with poorer mediolateral stability were more prone to injuries in the subsequent training period.  
Domain: Medical

**Example 3:**  
Abstract: 1. Angiotensin (Ang) II plays a major role in vascular remodelling. Matrix metalloproteinases (MMPs) and their tissue inhibitors (TIMPs) are involved in the tissue remodelling processes. The aim of the present study was to investigate whether AngII modulates TIMP-2 expression in rat aortic smooth muscle cells in vivo. 2. Angiotensin II (200 ng/kg per min, s.c.) or AngII + losartan (10 mg/kg per day, s.c.) or normal saline was administered continuously by osmotic minipumps to Sprague-Dawley rats for 1 week. In addition, the effect of endogenous AngII on TIMP-2 expression was evaluated in renovascular hypertensive rats (two kidney, one clip (2KIC) and one kidney, one clip (1KIC) models). Control rats (sham 2KIC and sham 1KIC rats) underwent sham-clipping of the left renal artery. At the end of the treatment, plasma renin activity was measured by radiimmunoassay. Systolic blood pressure (SBP) was measured twice a week by the tail-cuff method. 3. Exogenous AngII administration produced the expected increase in SBP ( $P = 0.02$ ) compared with the control saline-treated group. The increase in SBP was abolished in AngII + losartan-treated rats. Administration of AngII caused a significant increase in TIMP-2 expression ( $P = 0.01$ ) in rat aortic smooth muscle cells that was abolished in AngII + losartan-treated rats. In renovascular hypertensive rats, SBP was higher ( $P < 0.0001$ ) in 2KIC and 1KIC rats compared with the corresponding sham-operated rats. Plasma renin activity was higher ( $P < 0.01$ ) in 2KIC rats compared with the other groups. The expression of TIMP-2 was significantly ( $P < 0.05$ ) increased only in 2KIC rats. Our *in vivo* data demonstrate that exogenous and endogenous AngII increases TIMP-2 expression in rat aortic smooth muscle cells. This effect is not dependent on the AngII-induced increase in blood pressure and is mediated by angiotensin AT1 receptors.  
Domain: Biochemistry

**Example 4:**  
Abstract: 1. How and at what spatial scale(s) animals change their movements in relation to their environment is central to several topics in ecology and conservation, including foraging ecology, habitat selection and dispersal. A method (first-passage time analysis, FPT) has recently been proposed to measure changes in movements through the landscape, as an index of search effort at the pertinent spatial scales. This method seems largely applicable to an increasing number of studies using studies using satellite, radio-tracking or global positioning system (GPS), but its limits have not yet been assessed. 2. Here I used several movement simulations to examine the ability of FPT analysis to detect area-restricted search (ARS) according to different changes in movement, different patch structures and tracking accuracy. 3. FPT analysis was able to detect changes in movements when both speed and sinusity changed, or when the animal reacted to patch boundaries. It was also able to detect ARS within the same path at several spatial scales in patches (nested or not) of different sizes. 4. Tracking accuracy affected the detection of ARS by FPT analysis. With the widely used Argos system, a minimum of 13 locations in effective ARS was necessary to detect this behaviour; seven when velocity filtering was applied. Similarly, spatial error in location affected the estimation of the ARS scale value, but the application of velocity filtering reduced this effect. 5. Comparisons between a real GPS track and pathways simulating the Argos error showed that the time-sampling rate of locations (due to satellite-pass frequency) decreased the probability of detecting ARS at small scales (50% across the whole range of scales). A velocity filter enabled significant reductions in this effect. 6. Synthesis and application. Within limits, FPT analysis is highly suitable for animal movement analysis, either to quantify habitat use, or to determine the scale most relevant for describing an ecological system or factors affecting movement decisions. In anticipation of increasing applications of FPT analysis in applied ecology, I provide recommendations for the use of the technique with several tracking methods.  
Domain: CFC

ASSISTANT	<p>Abstract 1: An investigation was conducted on the genetic association of dopamine receptor-D4 and dopamine transporter genes with attention deficit hyperactivity disorder (ADHD) in children. The study found a significant association of specific gene variants with ADHD in the study population. Domain: Psychology</p>
Abstract 2:	<p>A study analyzed stabilometry in athletes to determine if injured athletes exhibit different stabilometric values before injury compared to non-injured athletes in different training periods. The results showed that athletes with poorer stabilometric values were more prone to sports injuries in subsequent training periods. Domain: Medical</p>
Abstract 3:	<p>Research focused on the modulation of tissue inhibitor of metalloproteinase-2 (TIMP-2) expression in rat aortic smooth muscle cells by angiotensin II (AngII) in vivo. The study showed that both exogenous and endogenous AngII can increase TIMP-2 expression, independent of blood pressure changes, in rat aortic smooth muscle cells. Domain: Biochemistry</p>
Abstract 4:	<p>The study evaluated the ability of first-passage time analysis (FPT) to detect changes in animal movements in relation to the environment at different spatial scales. The research utilized movement simulations to assess the accuracy of FPT analysis in detecting area-restricted search behavior, depending</p>
USER	<p>Now I will start providing the abstracts and classify them into one of the 7 domains</p>
ASSISTANT	<p>Sure! Please go ahead and provide the abstracts for classification.</p>

<b>USER</b>	<p>XML is a pervasive technology for representing and accessing semi-structured data. XPath is the standard language for navigational queries on XML documents and there is a growing demand for its efficient processing. In order to increase the efficiency in executing our navigation XPath query processor, namely descendants, ancestors, children and parent, we introduce a tree-based approach. The main idea is to build a tree structure for each node in the XML document. This tree structure is used to store the information about the descendants, ancestors and parents of each node. We also propose a new set of operations which will directly return the desired sub-trees, avoiding to create a passing through nodes and edges. Our solution is based on the well-known SeL4 microkernel. The experiments show that our solution is competitive with the state-of-the-art solutions.</p>
<b>DEVELOPER</b>	<p>We present a novel approach for the implementation of the NESTOR system. The NESTOR system is a framework for the development of distributed systems by using several mobile devices (e.g. collections), NEON 2000 (Wi-Fi collection), and live video (use standard mobile devices (SmartPhone and Tablets)). We show that NESTOR based distributed system can be used for the development of distributed systems and the low computational overhead of both memory occupation and pre-processing time. (C) 2010 Elsevier Ltd. All rights reserved.</p>
<b>ADMINISTRATOR</b>	<p>Domain: CS</p>
<b>USER</b>	<p>The integration of renewable energy sources increases the complexity in maintaining the power grid. In particular, the highly dynamic nature of generation and consumption demands for a better utilization of energy resources, which can be used for storage infrastructure, can only be achieved through demand-response. Accordingly, the availability of energy and potential consumption must be balanced in real-time. In this paper, we propose a framework for the real-time exchange of information between the grid and its consumers. In this paper, we investigate the capability of an intelligent energy management system to timely inform users about usage in case of emergency. The proposed system consists of two parts: a central server and a local controller. The central server performs the global optimization of the energy system, while the local controller performs the design of more effective systems. The outcome is the formulation of energy efficiency policies for residential buildings, as well as the design of an energy management system, consisting of hardware and software components. This paper presents the proposed framework, which can reuse the open source software, called EnergyOS, as a platform for the implementation of the concepts. The system can be implemented as an ad-hoc wireless network, displaying disaggregated and aggregated data. The ad-hoc wireless network can be used to automatically analyze usage and provide timely feedback. The ad-hoc wireless network is a potential of lots of savings using renewable energy sources.</p>
<b>ASSISTANT</b>	<p>Domain: MAE</p>
<b>USER</b>	<p>The hybrid electric vehicle engine with the operating strategy generated by the computer, relevant thermal losses, for balanced mechanical system, possibility of four-stroke operation of the engine, possibility of the optimal operation of the engine, optimal torque circular of the electrical generator, etc. Because the engine can optimally operate over wide range of the output power. It enables to improve electrical system of the hybrid vehicle by usage of supercapacitors, and optimized energy flows, in this way the competitive hybrid electric vehicle can be created.</p>
<b>ASSISTANT</b>	<p>Domain: ECE</p>

**Classification accuracy:** In this case the classification accuracy managed to reach 0.6 (60%).

(Optional) Observations: Unlike previous cases where there were complexities handling certain domain classifications in this case the domains were more, or less appropriately classified, and even when there was incorrect classification it was observed that the model was quick to realize when prompted, and accurately changed it to the correct format.

Part II – Fine-tuning a pre-trained encoder-only model

For this part, I ask that you return a Google Colab notebook with all the necessary code to answer these questions. A template is provided. Note that this template uses pytorch. If you'd rather use TensorFlow, please feel free to make the required modifications. **In general, the colab template is provided as an aid but it is not required to follow it. The only requirements**

are that the notebook that you submit is clear (i.e., easy to read), runs without errors, and doesn't require paying Colab resources to run.

#### Question #4

The task is to fine-tune [Distilbert](#) on the training data by adding an additional classification layer on top of it. Please refer to the [Google Colab](#) for a skeleton of how to perform this task. Here are the steps to follow:

- 1) Prepare the datasets and load the pre-trained DistilBert model (loading code is provided)
- 2) Implement the DataLoader: see documentation [here](#)
- 3) Implement the model: this model has a base of DistilBert and you should add the necessary layer(s) to train it for classification
- 4) Initialize model, loss, and optimizer (code is provided and only requires certain parameters to be set). Code to free DistilBert parameters is also provided.
- 5) Implement the learning loop (template is provided, lots of tutorials available online, for instance:  
[https://colab.research.google.com/github/rbkhb/NLP\\_IMC/blob/master/BERT\\_Fine\\_Tuning\\_Sentence\\_Classification.ipynb](https://colab.research.google.com/github/rbkhb/NLP_IMC/blob/master/BERT_Fine_Tuning_Sentence_Classification.ipynb))
- 6) Evaluate fine-tuned model on the test set (code provided)

This part can be run on cpu or T4 GPU.

**Classification accuracy:** In this case with the frozen Distilbert model the classification accuracy was 0.2 (20%).

(Optional) Observations: The accuracy generated by the frozen Distilbert model with fine tuning was far lower than the accuracy generated using the gpt 3.5 model with, or without prompt engineering. This difference can be attributed to the fact that there wasn't sufficient data available for fine tuning a large language model resulting in overfitting of the model causing the lowered accuracy.

#### Question #5

The task is to fine-tune [Distilbert](#) on the training data by adding an additional classification layer on top of it but without freezing the Distilbert model. In other words, we fine-tune the entire model end-to-end. Please refer to the [Google Colab](#) for a skeleton of how to perform this task. Here are the steps to follow:

- 1) Prepare the datasets and load the pre-trained DistilBert model (loading code is provided)
- 2) Implement the DataLoader: see documentation [here](#)

- 3) Implement the model: this model has a base of DistilBert and you should add the necessary layer(s) to train it for classification
- 4) Initialize model, loss, and optimizer (code is provided and only requires certain parameters to be set)
- 5) Implement the learning loop (template is provided, lots of tutorials available online, for instance:  
[https://colab.research.google.com/github/rbkhb/NLP\\_IMC/blob/master/BERT\\_Fine\\_Tuning\\_Sentence\\_Classification.ipynb](https://colab.research.google.com/github/rbkhb/NLP_IMC/blob/master/BERT_Fine_Tuning_Sentence_Classification.ipynb))
- 6) Evaluate fine-tuned model on the test set (code provided)

This part can be run on a T4 GPU.

**Classification accuracy:** In this case with the unfrozen Distilbert model the classification accuracy was 0.15 (15%).

**(Optional) Observations:** The accuracy with unfrozen layers in the model was lower than with frozen layers, and this accuracy is still far below the accuracy obtained from gpt 3.5 model indicating that still due to a lack of data for fine tuning the model there is overfitting observed with the model resulting in lower test accuracy.