

PROMPTING PRE-TRAINED MODELS VS TRADITIONAL MACHINE LEARNING

In recent years, the words artificial intelligence[1] and machine learning (ML) became prevalent.

This technology has revolutionized how we approach problem-solving across various industries, from healthcare to finance, and even in our daily interactions with technology whether at work[2] or in our personal lives.

Recently, with the advent of GPT a new paradigm has appeared: from traditional ML models, meticulously crafted and trained for specific tasks, we are now shifting to pre-trained models that can be adapted for a multitude of applications through prompt engineering. Understanding these two approaches is key to unlocking their potential in today's data-driven world.

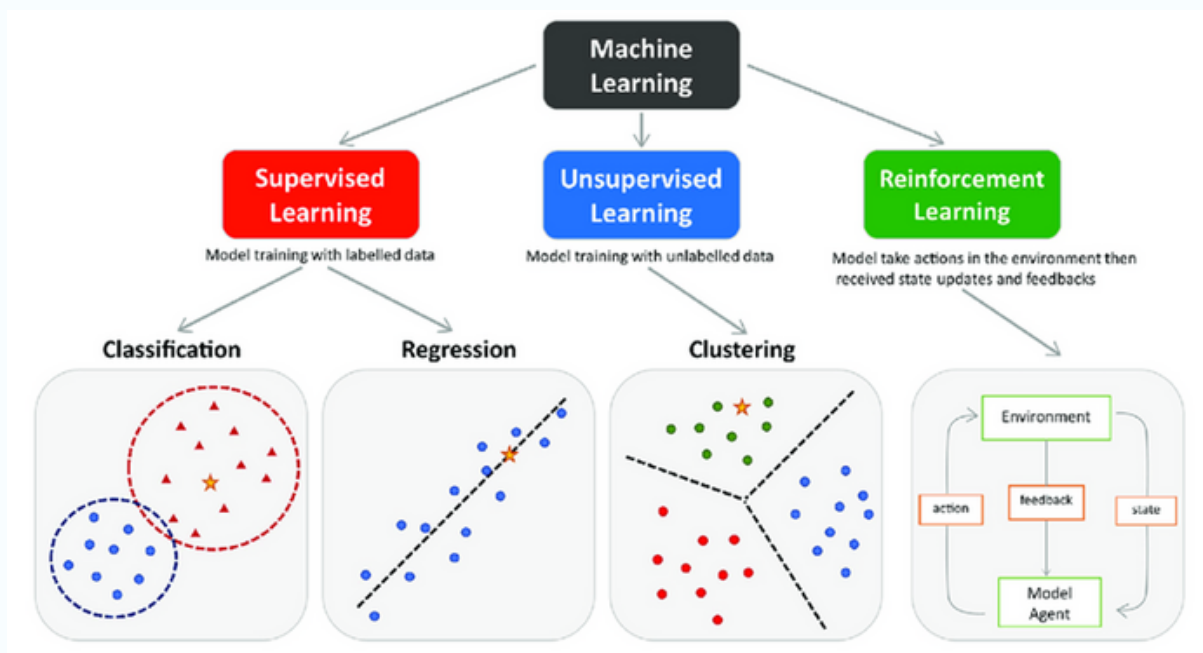


What is Traditional Machine Learning?

Traditional ML is a method where algorithms learn from data to make predictions or decisions. At its core, this process involves six primary steps:

- **Data collection**, where data is harvested from various sources.
- **Data pre-processing**, where raw data is cleaned and structured.
- **Data annotation**, where data is given labels to be predicted.
- **Feature engineering**, where the most relevant aspects of the data are identified and highlighted.
- **Model training**, where the algorithm learns from this prepared data to make future predictions.
- **Model evaluation**, where the model is evaluated on new data to validate its performance.

Figure 1: Different kinds of ML Techniques[1]



For example, if the task involves translating text from English to French, the process would unfold as follows: First, we need to gather a bilingual corpus in both languages. Then we can do some pre-processing, such as removing links. Then, we annotate, which in this case means pairing each English sentence with its corresponding known French translation. After this, we can engage in feature engineering, like encoding the words using word embeddings[3]. Finally, we train the model and evaluate its performance.

The strength of traditional ML lies in its specificity. Models are tailor-made for the task at hand, whether it's predicting stock market trends or diagnosing diseases from medical images. This customization requires a deep understanding of both the problem and the data, necessitating significant investment in data collection and processing, as well as model development and tuning.

What is Prompt Engineering with Pretrained Models ?

Pretraining is a specific ML approach that has emerged with the rise of large of models like GPT-3, BERT, and others[4]. These models have been trained on vast datasets, encompassing a wide range of knowledge and linguistic structures. Instead of building a model from scratch as in the traditional approach, we can use pre-trained models that then act as a knowledge base on which prompt engineering can be performed.



Prompt engineering[5] involves crafting inputs (prompts) to these pre-trained models in a way that guides them to produce the desired output. Thus, the model training step is replaced by prompt tuning where the difficulty lies not in tuning hyperparameters or costly training computation but rather in meticulously defining the input prompt with the right examples, context, and directives. This task requires a deep knowledge of the inner workings of the pre-trained model and the problem to be solved to ensure that the examples used in the prompt are as representative and diversified as possible to represent the closest possible to the real distribution of the input-output pairs that define the problem.

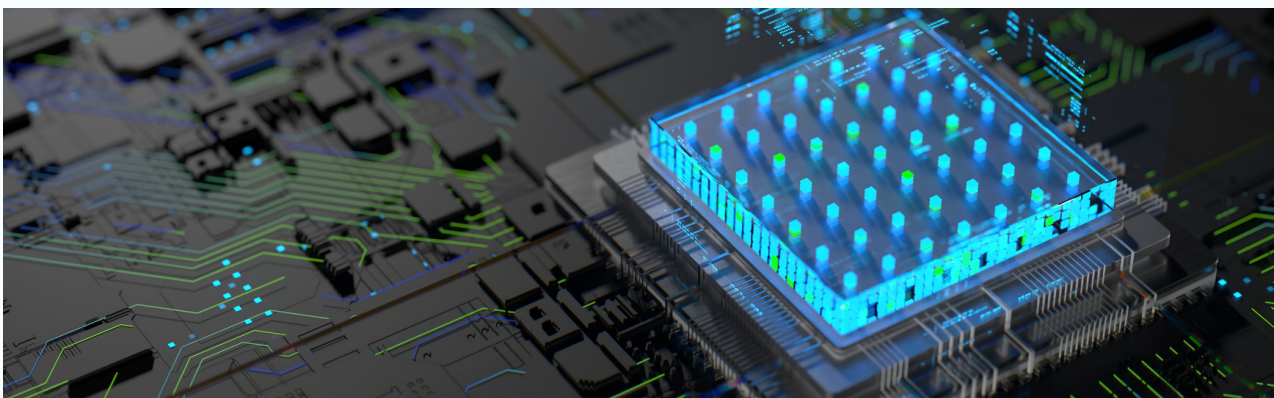
In this case, the task of translating text from English to French is much simpler to accomplish. All we need to do is equip ourselves with a pre-trained model such as GPT-4 and provide it with a prompt. This prompt contains the definition of the task, a few example sentences along with their translations, and finally, the current sentence to be translated, as can be seen in the image above. And that's all! GPT-4, having already been pre-trained on a huge corpus, has already learned the concept of translation; it just needs to be asked correctly with a prompt.

This approach leverages the extensive learning these models have already undergone, making it possible to apply them to a variety of tasks without the need for extensive dataset curation or model training specific to each task; we call this transfer learning. From generating creative content to answering complex queries, the versatility of prompt engineering with pre-trained models is shocking.

Through this approach, the burden of data collection, preprocessing, and annotation is greatly reduced and, in some cases, eliminated. This aspect is crucial to acknowledge, as these steps typically constitute a significant portion of the modeling task in terms of time and cost. The use of pre-trained models substantially alleviates these demands, streamlining the process of deploying ML solutions.



Figure 2: Example of a prompt



Pros and Cons

Traditional Machine Learning

Pros : Traditional ML is highly Customizable as it can be tailored to specific tasks with a high degree of precision. Moreover, we collect and preprocess the data ourselves which brings more control and validation over what the model ingested.

Cons : However, training a model is a resource-intensive process that requires significant investments. Data collection, preprocessing, and annotation are costly and generally account for the majority of the time taken in ML projects. Moreover, training a model can also be extremely costly as well as time-consuming.

Prompt Engineering on Pre-trained Models

Pros : Prompt Engineering on Pretrained Models limits the need to deal with data which leads to cheaper and faster model deployment. It also means that such models can quickly adapt to changes in the problem specification.

Cons : The main problem with pre-trained models is that we have little oversight on the data sources and preprocessing methods used for their creation. Thus, for sensitive tasks such as medical or financial problems, such models may not be appropriate. Nonetheless, we can also use pre-trained models on medical or financial data to alleviate this issue.

Conclusion

In the futur, Prompt Engineering on Pretrained Models could become a ubiquitous in industrial applications considering its speed of deployment and reduced cost. We might witness the emergence of more sophisticated prompt engineering techniques, making pretrained models even more versatile. Simultaneously, traditional ML will continue to advance, especially in areas requiring extreme precision and data oversight.

In conclusion, both traditional ML and prompt engineering with pre-trained models offer unique advantages and drawbacks. The choice between them should be guided by the specific requirements of the task at hand, considering factors like data availability, the need for customization, and resource constraints. As the field evolves, the synergy of these two approaches may pave the way for more innovative and efficient solutions in the realm of artificial intelligence.

Bibliographie

- 1.Poumay (Ph.D.) J. Comprendre l'intelligence artificielle. The Thought Process. Published February 10, 2022. Accessed March 11, 2024. <https://thethoughtprocess.xyz/science/informatique/2022/02/10/comprendre-lintelligence-artificielle/>
- 2.Poumay (Ph.D.) J. L'intelligence artificielle : ses développements passés et son impact futur sur le travail. The Thought Process. Published September 21, 2023. Accessed March 11, 2024. <https://thethoughtprocess.xyz/non-classe-fr/2023/09/21/lintelligence-artificielle-ses-developpements-passes-et-son-impact-futur-sur-le-travail/>
- 3.Poumay (Ph.D.) J. Embeddings : lier les mots aux nombres. The Thought Process. Published June 19, 2020. Accessed March 11, 2024. <https://thethoughtprocess.xyz/science/informatique/2020/06/19/embeddings-lier-les-mots-aux-nombres/>
- 4.Poumay (Ph.D.) J. Notes de lecture #6 : Modèles de langage. The Thought Process. Published August 18, 2022. Accessed March 11, 2024. <https://thethoughtprocess.xyz/science/informatique/2022/08/18/notes-de-lecture-6-modeles-de-langage/>
- 5.Poumay (Ph.D.) J. Le Prompt Engineering : Comment parler avec les IA. The Thought Process. Published November 22, 2023. Accessed March 11, 2024. <https://thethoughtprocess.xyz/non-classe-fr/2023/11/22/le-prompt-engineering-comment-parler-aux-ia-comme-chatgpt/>