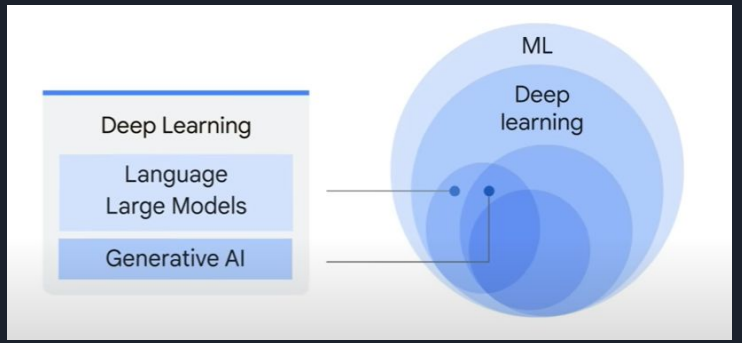
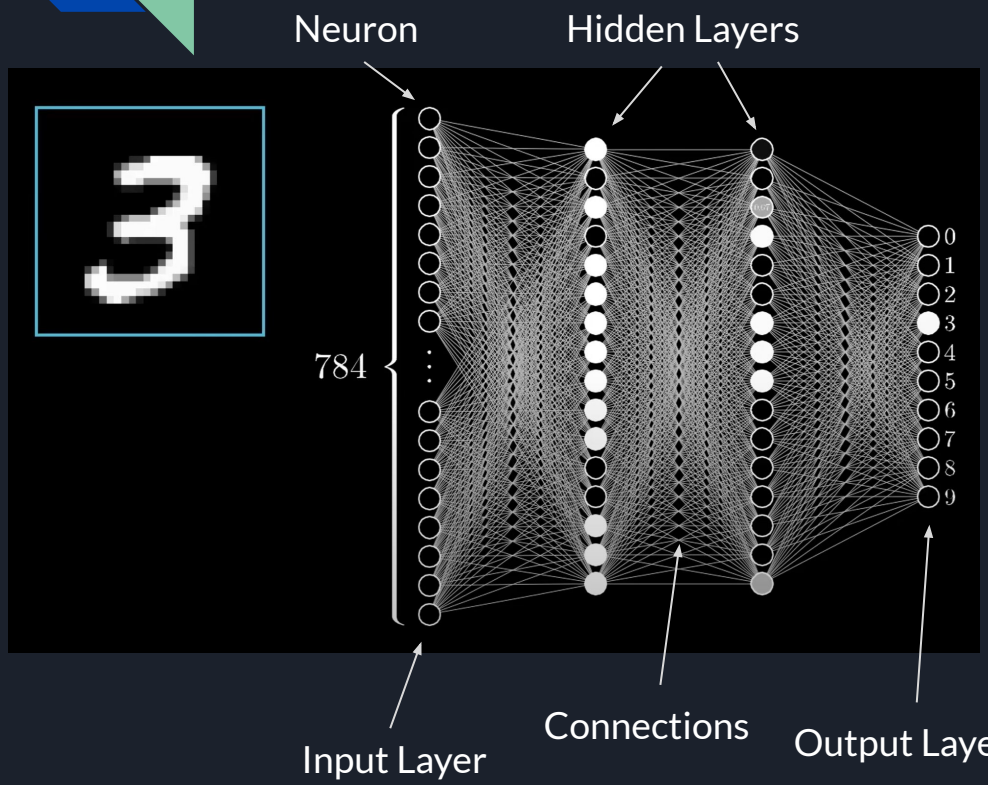
Large Language Models (LLMs)

Overview - What Are Large Language Models?

- “Language Model” - An artificial intelligence system designed to understand, generate, and interact with human language at a vast scale
- “Large” - Referring to the massive amount of data used to train such models, sometimes hundreds of terabytes or petabytes in size
- Essentially neural network models that are designed to understand human language and word context



Overview - What Are Neural Networks?



Neuron: Receives input from other neurons or the external world and computes a weighted sum of inputs

Connections: Adjustable parameters determining connection strength (weight and bias)

Input Layer: Receives external data as the initial input

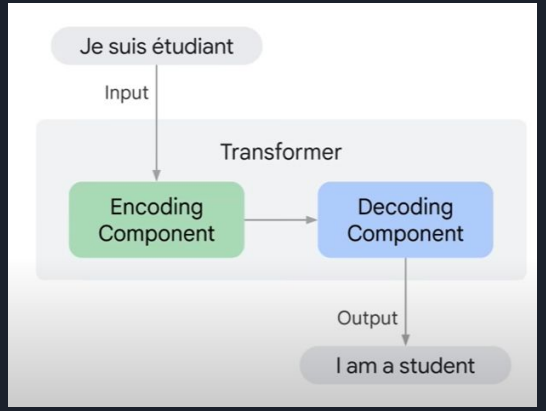
Hidden Layers: Multiple layers that process and extract data features

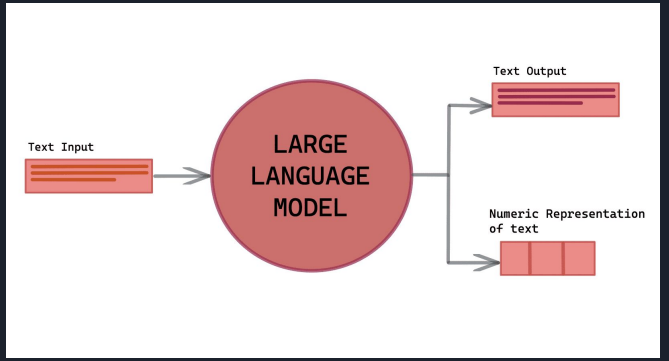
Output Layer: Provides final predictions or outcomes

Optimization Algorithm: Updates weights and biases to minimize errors (Function cost/Loss Function & Backpropagation)

Overview - How Do LLMs Work?

- Transformers pre-process text as numerical representations through the encoder to understand the context of words and phrases with similar meanings
- During training, the model learns to predict the next word in a sentence using the context learned from previous sentences
    - Ex: “The sky is…” or “Berlin is to Germany as Paris is to…”
- Gradually improves its word predictions until it can reliably generate coherent sentences





Examples Of LLMs

Chat GPT

Generative **P**re-training **T**ransformer

BERT

Bidirectional **E**ncoder **R**epresentations from **T**ransformers

Flan T5

Fine-tuned **LA**nguage **N**et + **T**ext-**t**o-**T**ext **T**ransfer **T**ransformer

Strengths And Weaknesses

Strengths

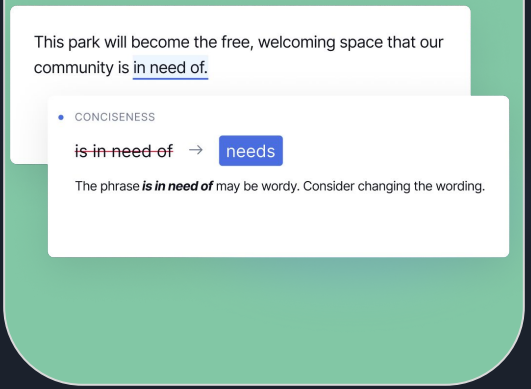
- Near-human performance: LLMs can perform well on core NLP tasks, such as machine translation, natural language generation, question answering, and image captioning.
- Real-time adaptation: LLMs can be quickly retrained on new data.
- More accurate language representation: LLMs can provide a more accurate representation of language.
- Probabilistic: LLMs are probabilistic and can work with natural language and its ambiguities.

Weaknesses

- Difficult to interpret: LLMs are often based on complex algorithms that can be difficult to understand.
- Data limitations: Expansive amounts of data required to train
- Biased: LLMs can be biased towards certain groups of people (whoever trained/created the model)
- Hallucination: LLMs can produce high quality text which contains factually incorrect information.

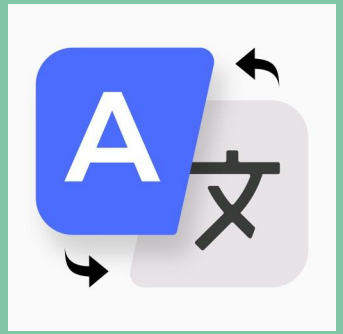
Conclusion - Can it benefit the project?

Text Correction and Formatting



Contextual Understanding

Language Translation



Sources

https://aws.amazon.com/what-is/large-language-model/

https://www.elastic.co/what-is/large-language-models#benefits-of-large-language-models

https://www.youtube.com/watch?v=zizonToFXDs&ab_channel=GoogleCloudTech

https://www.youtube.com/watch?v=iR2O2GPbB0E&ab_channel=GoogleforDevelopers

https://www.youtube.com/watch?v=5sLYAQS9sWQ&ab_channel=IBMTechnology

https://aws.amazon.com/what-is/transformers-in-artificial-intelligence/#:~:text=Transformers%20are%20a%20type%20of,tracking%20relationships%20between%20sequence%20components.

https://www.techtarget.com/whatis/feature/12-of-the-best-large-language-models

https://link.springer.com/article/10.1007/s10849-023-09409-x