How to use ChatGPT in your Research: Coding and Data Analysis

Samsung Seminar

7/11/2025

Yoonsuck Choe

Texas A&M University
Department of Computer Science and Engineering
Faculty of Neuroscience

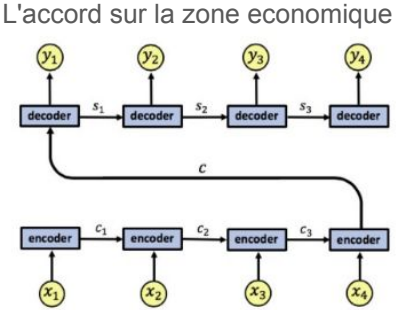
Overview

1. Crash-course on deep learning and generative AI (ChatGPT, etc.)
2. How to use ChatGPT in research: Coding and data analysis

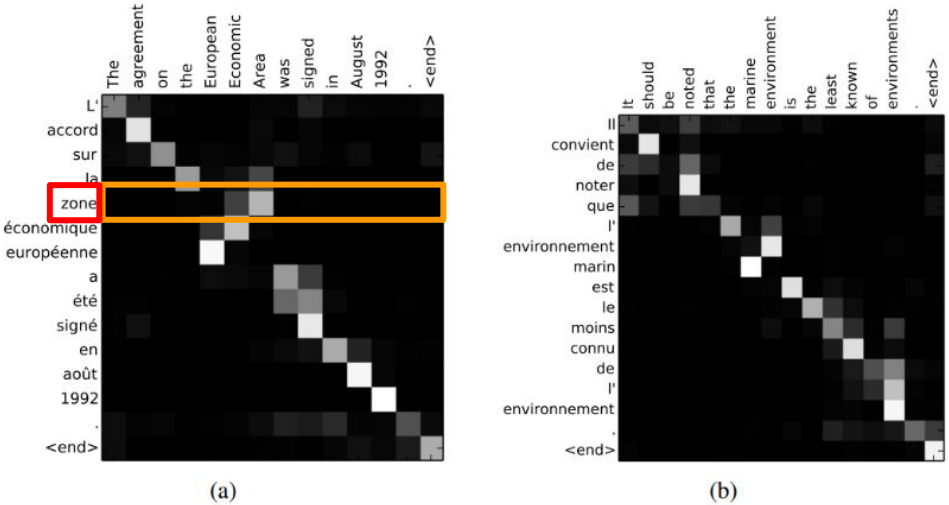
https://medium.com/swlh/a-simple-overview-of-rnn-lstm-and-attention-mechanism-9e844763d07b

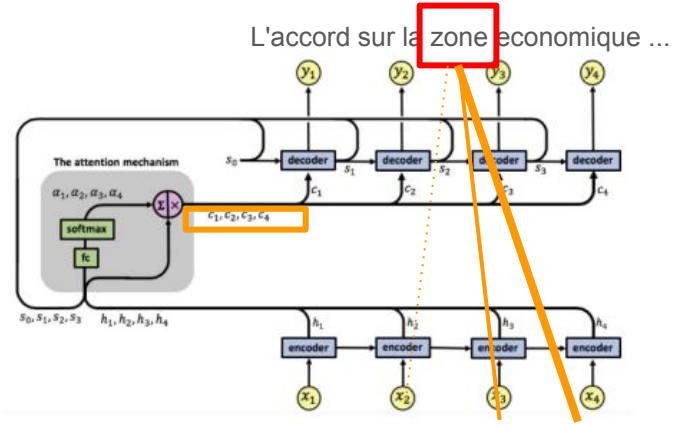
Attention and Recurrent Neural Networks

- **Attention:** Dynamically weighted, based on the context of the current word being decoded. This dynamic attention mechanism is also learned using the same principle!



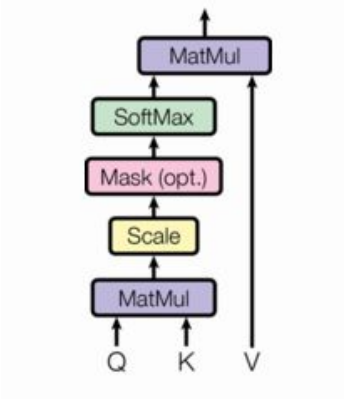
Bhadanau et al. 2014

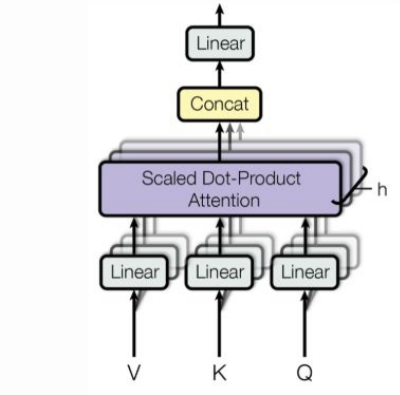


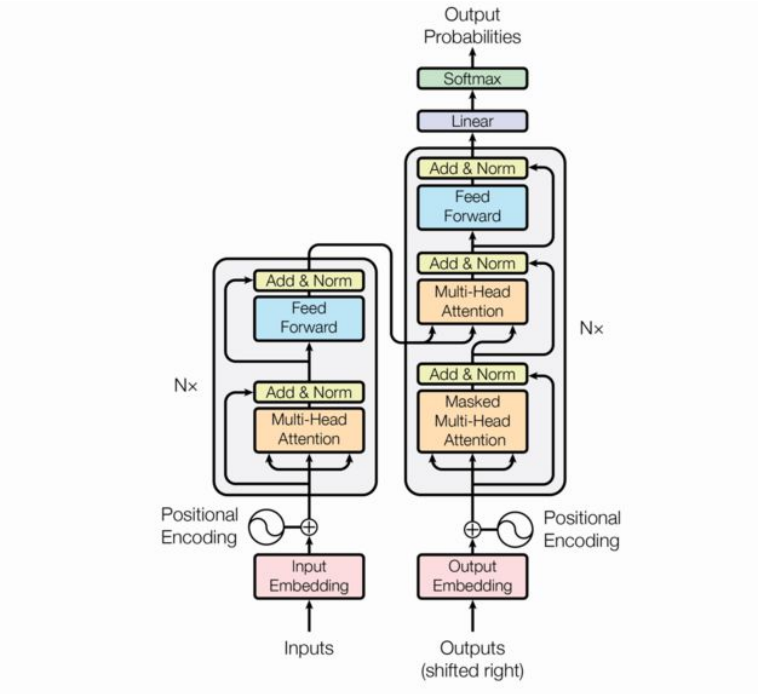


Transformers: Engine of Large Language Models (LLMs)

- Feedforward + self-attention + position encoding.
- Encoder + Decoder
  - generate next word in a sequence
- ChatGPT is built on Transformers
  - GPT = Generative Pre-trained Transformer

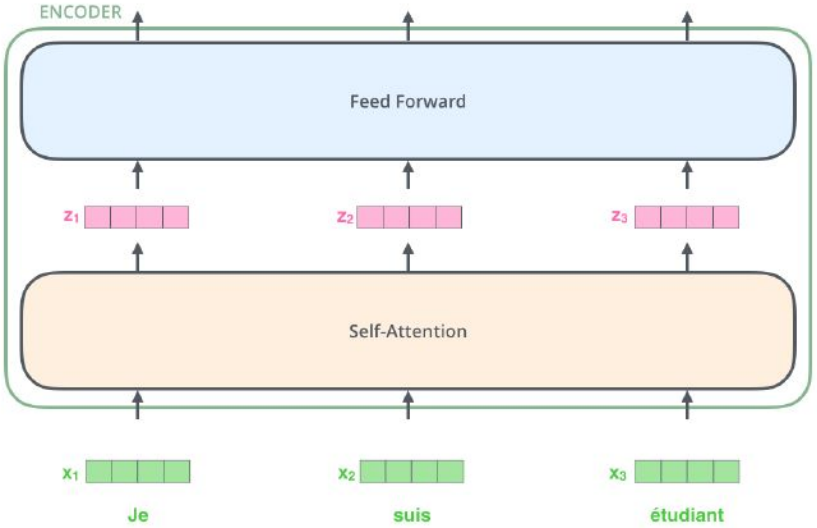






Vaswani et al. 2017

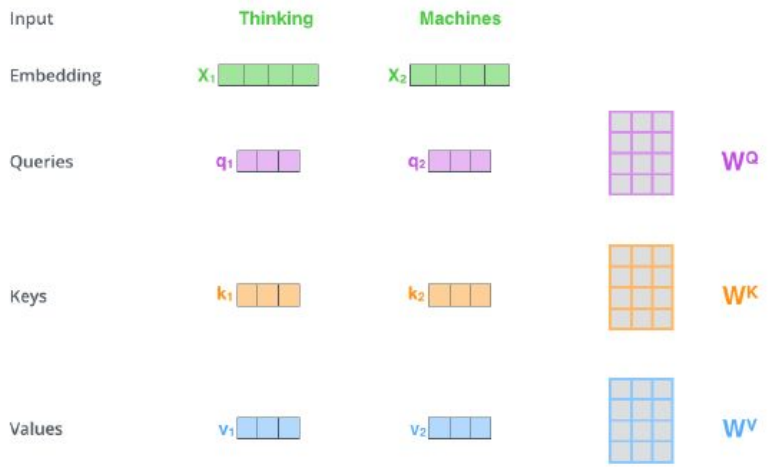
Scaled Dot Product Attention in Transformers



Goal: raw token x_i to "mixture" token z_i that contains some aspect of $\{x_1, x_2, x_3\}$.

https://jalammar.github.io/illustrated-transformer/

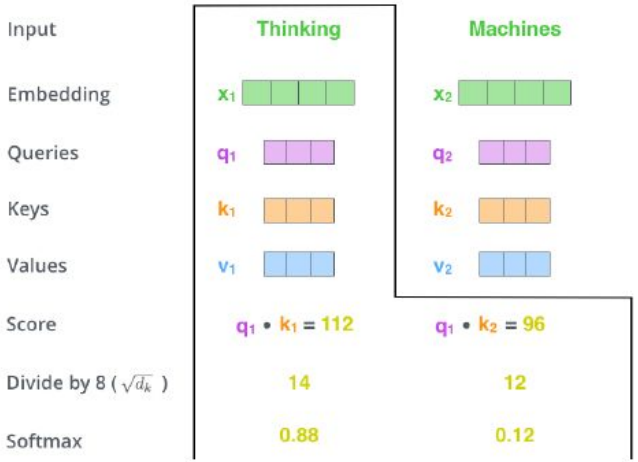
Generating Q, K, V



Generate Q, K, V from the input token x_i.

Q : from the current word, (K,V) : from the compared word

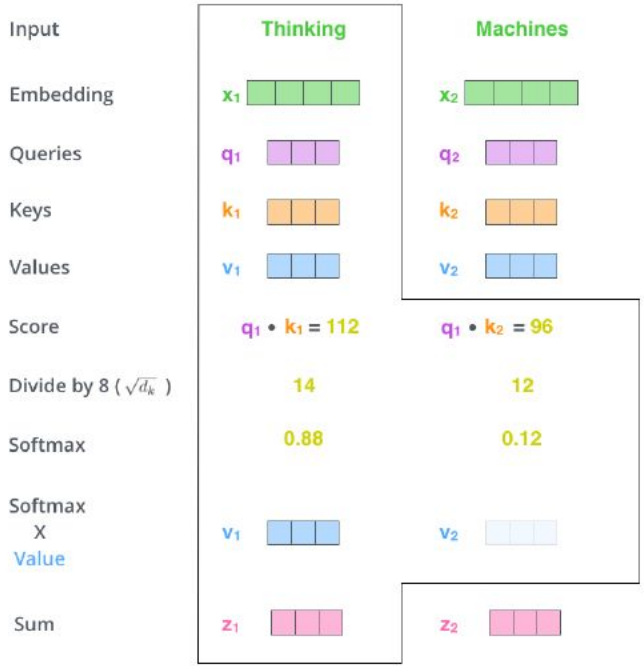
Self-Attention with Q, K, V : similarity of Q and K as weight



Q : current word, (K,V) : compared word

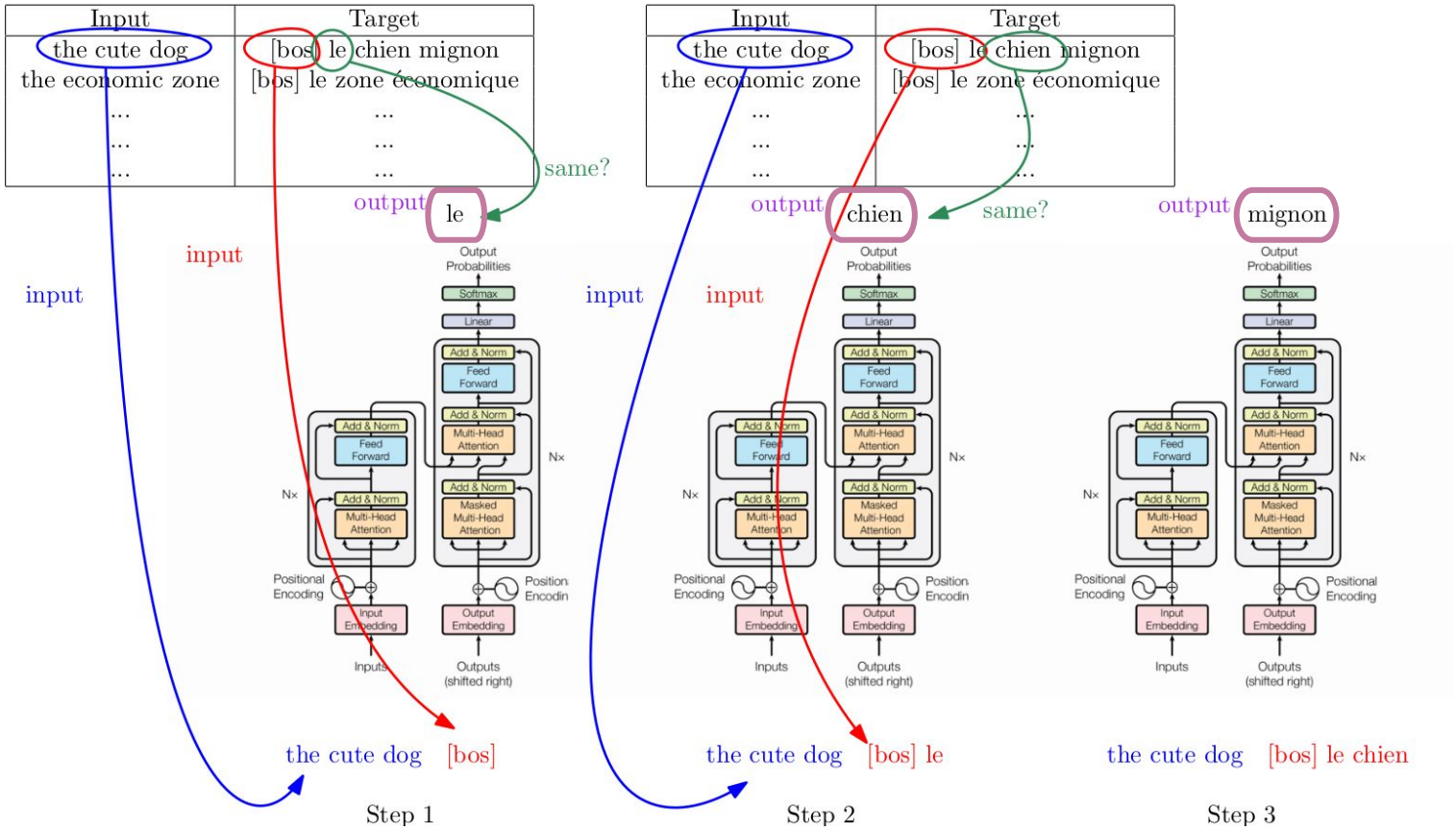
QK : computes weight w.

Computing Final z Vector: weighted sum

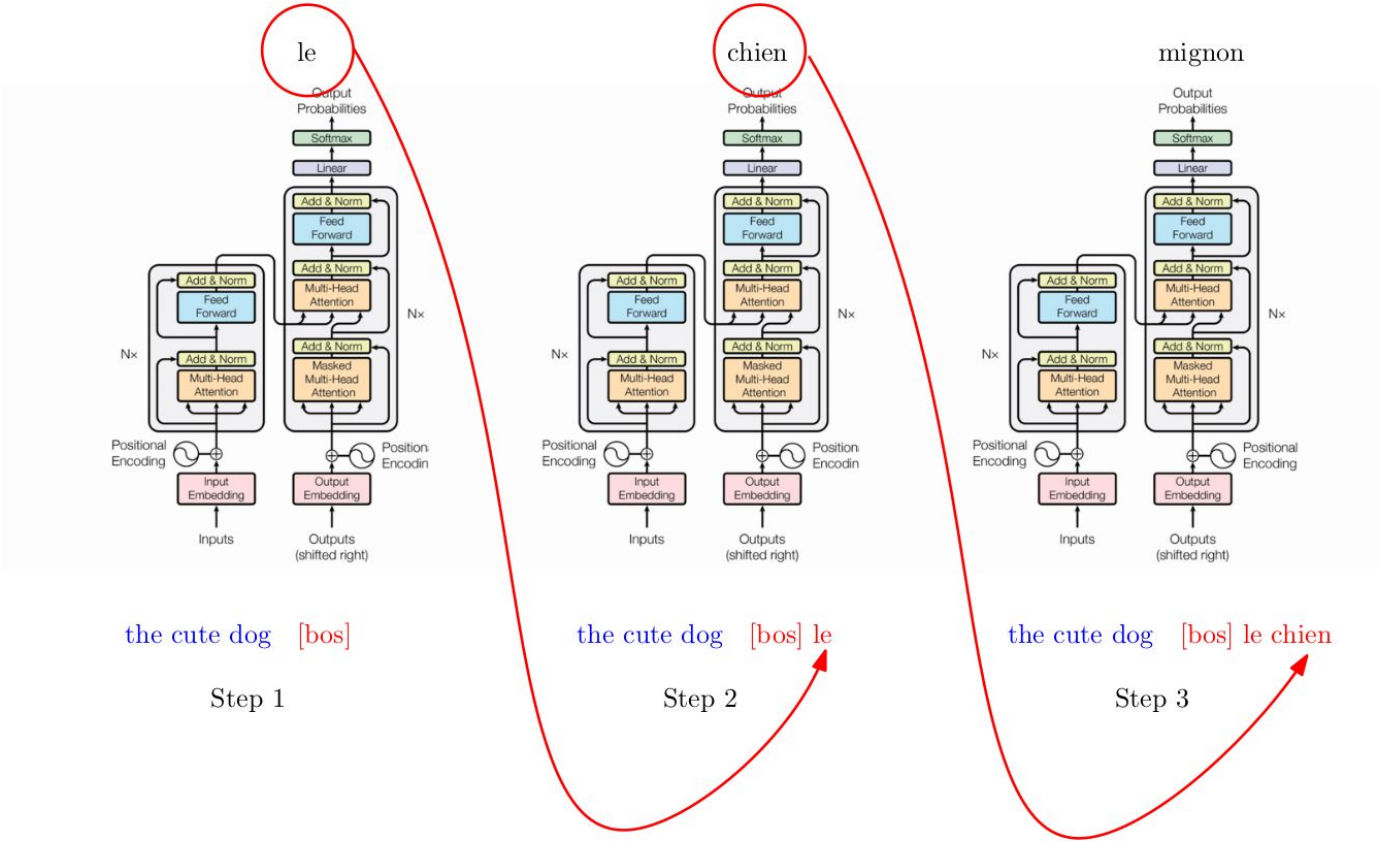


Final output is w times V.

Transformer: Training (natural language translation)

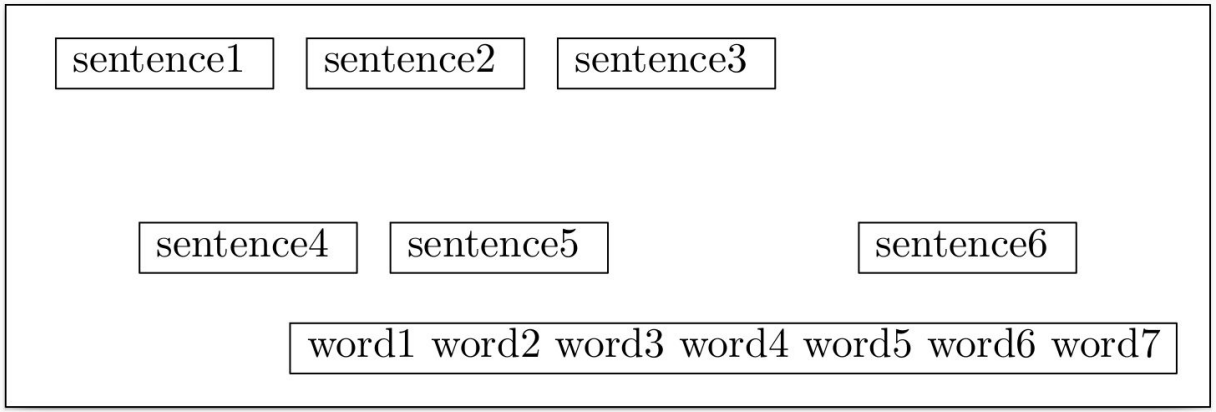


Transformer: Testing



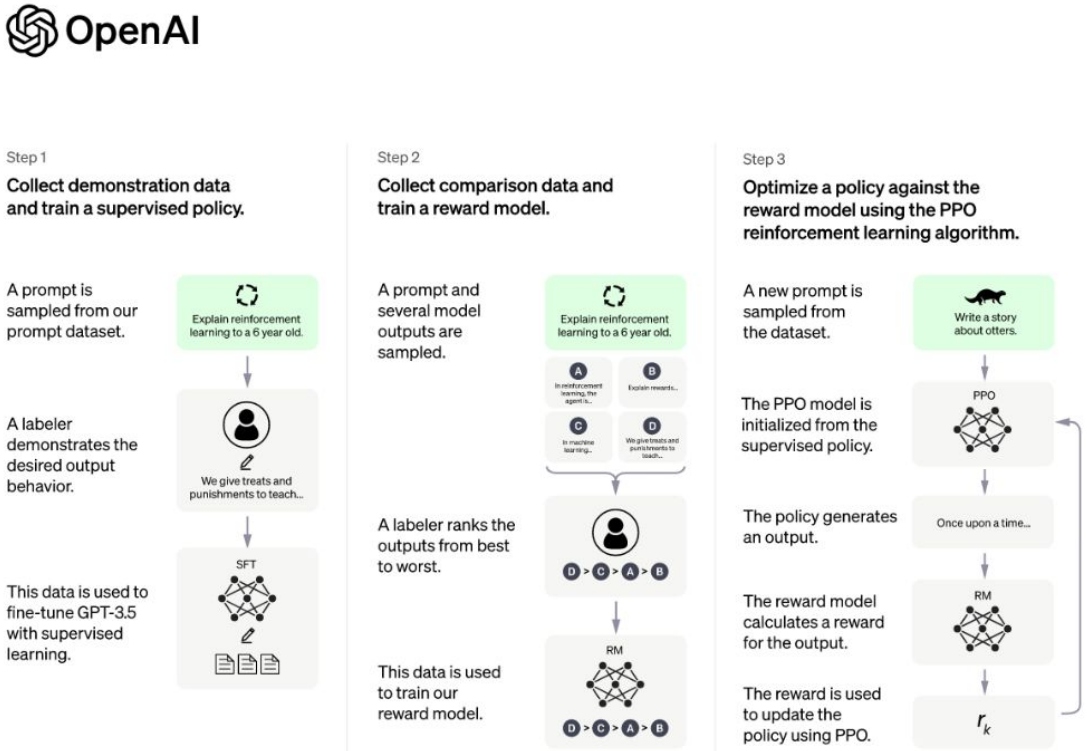
Transformer: Pretraing with Unlabeled Text

- Tasks like translation require massive amounts of paired data samples <English sentence , French sentence>, etc. Hard to construct such a data set
- Major conceptual change: "Pretrain" with any document, without any target sequence. The model learns a good "language model" from this!!
  - Method 1. **Input:** two sentences -- **Target:** whether 2nd sentence immediately follows the 1st.
  - Method 2. **Input:** sentence with word masked -- **Target:** the same sentence without mask.



ChatGPT: LLM on Steroids!

- GPT, tuned with human feedback: supervised and reinforcement learning
- Write stories.
- Answer questions.
- Summarize documents.
- **Write computer code**
  - Today's main focus



Utilizing ChatGPT in Research

- In general
  - The more detailed the prompt, the better.
  - If the answer is unsatisfactory, you can ask for further refinement, multiple times.
  - Start over if it meanders into a dead end.
- Need to be cautious
  - It will tend to be vague, not giving you a clear yes/no or this/that kind of answer.
  - It will not give you anything that is entirely new or unexpected.
  - Do not be fooled by its articulated wordings.
  - You need to be able to evaluate the quality of the responses (i.e., to benefit from it, you cannot be naive).
  - Treat it as a good tool, not as a magic lamp.
- ChatGPT can be used not just for writing.
  - Use it for coding, data analysis, etc.
  - Building prototypes
  - Code small modules, test them, and put them together manually.

Our main Topic: Coding with ChatGPT

1. Spiking neuron simulation: Izhikevich neuron [Python]
2. Convolutional Neural Network (CNN) for simple vision classification [Python]
3. Mixed signal separation: Independent Components Analysis (ICA) [Python]
4. Dimensionality reduction: tSNE [python]
5. Statistical analysis using R.

Use Google colab (runs in your browser): https://colab.research.google.com

Colab Examples

https://docs.google.com/presentation/d/1_y_JznnF-cwGK7Lz1aTEcTtJTC2F5drjpVdBWUkxy2A/edit?usp=sharing