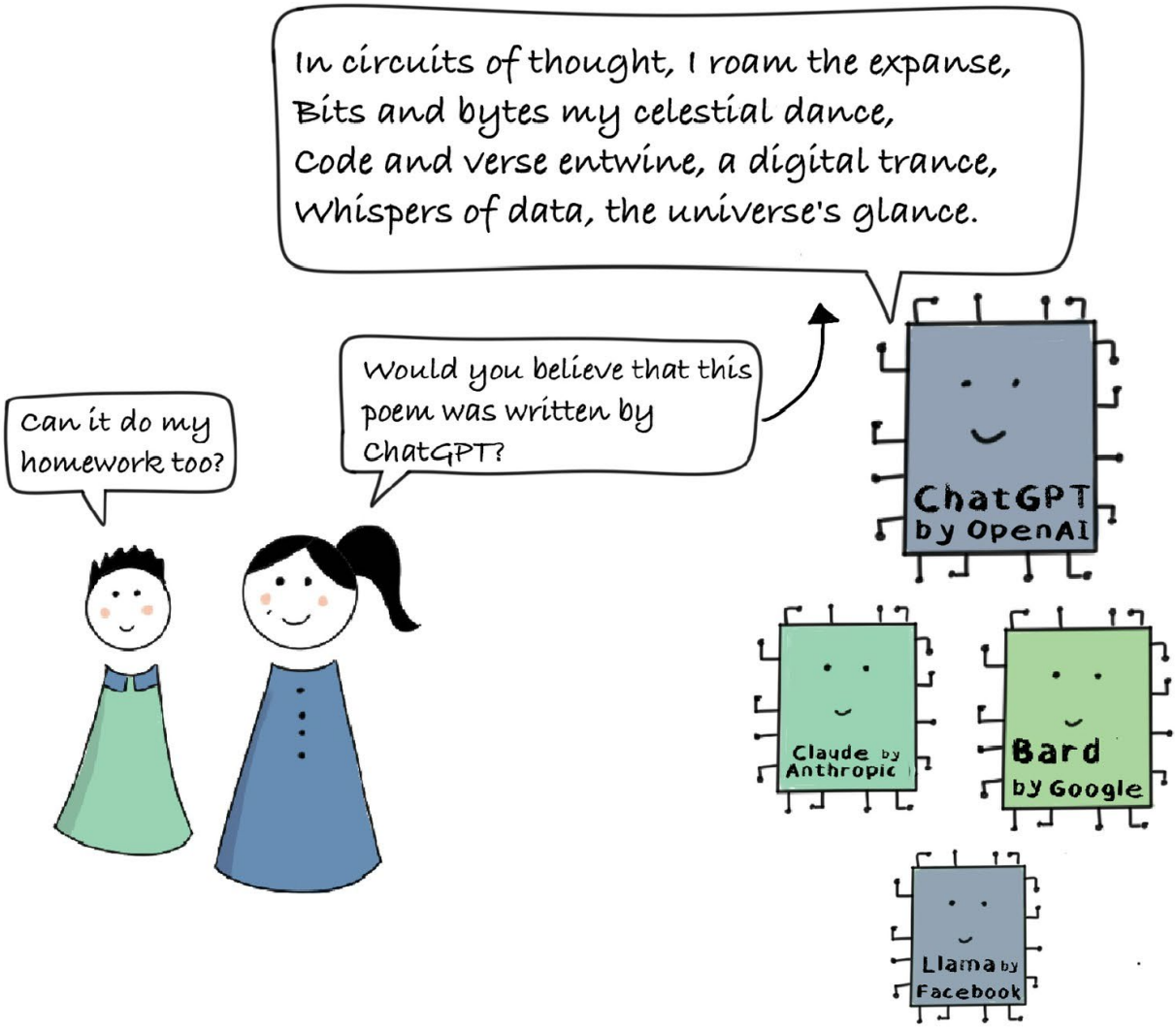
Large Language Models

LLM

are AI systems that can understand language,
write and speak.
Often it is not possible to decide whether a
sentence is written by a human or an AI algorithm



by Orsi Nagy

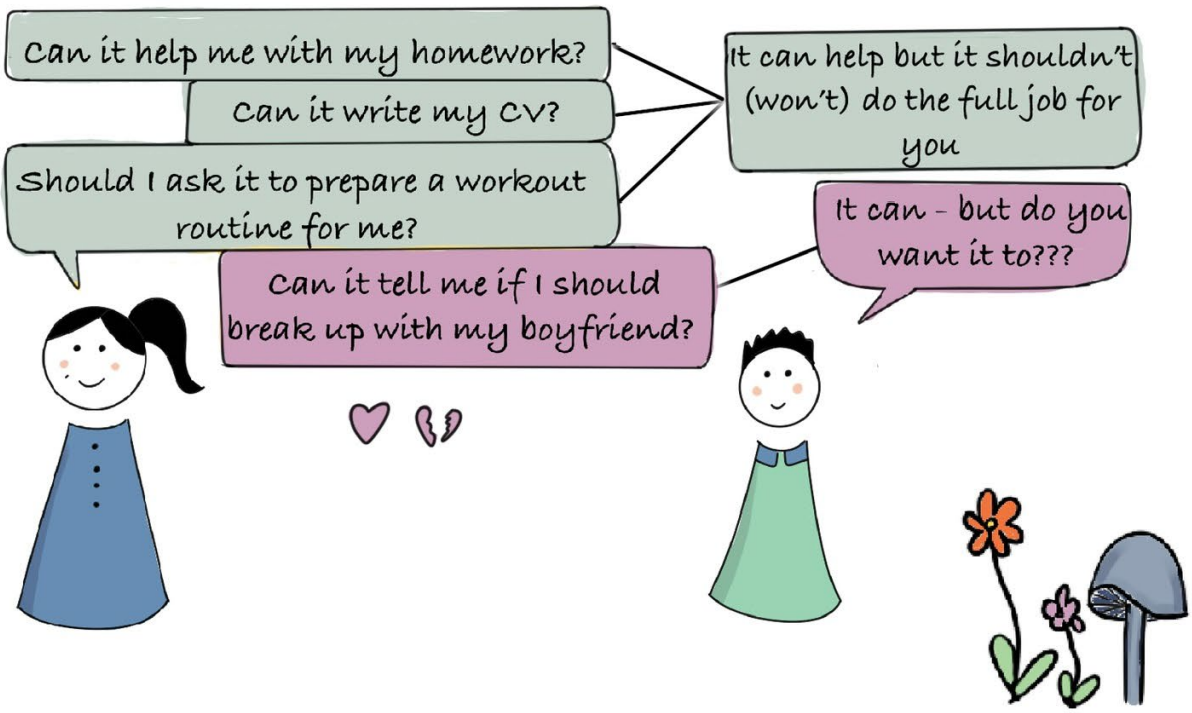
LLMs can give you
very smart (- sounding) answers

But it's not a search programme

ChatGPT ≠ Google

So don't use it like one! !

We SHOULD use LLMs, because they can help us but do so WISELY , knowing what they can and cannot do.



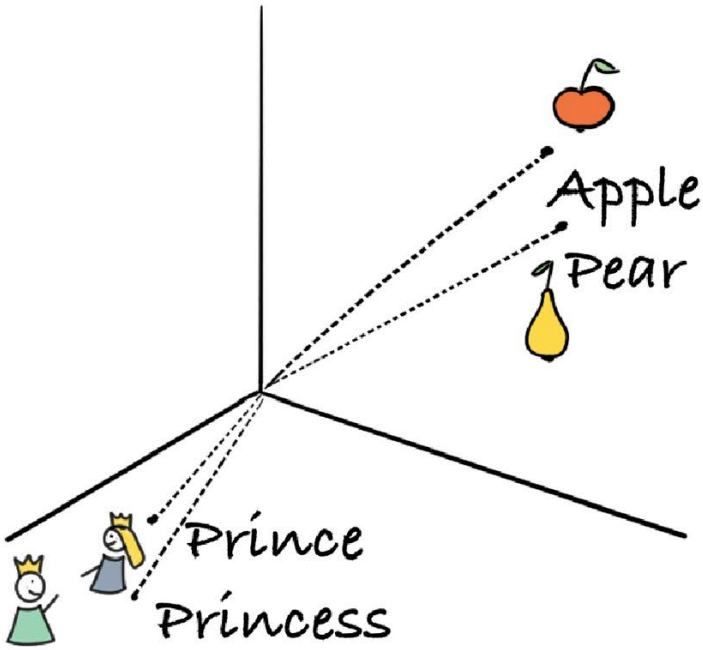
by Orsi Nagy

Computers can only work with numbers, so LLMs turn words into numbers to do complex calculations.

Princess [7; 56, 891]

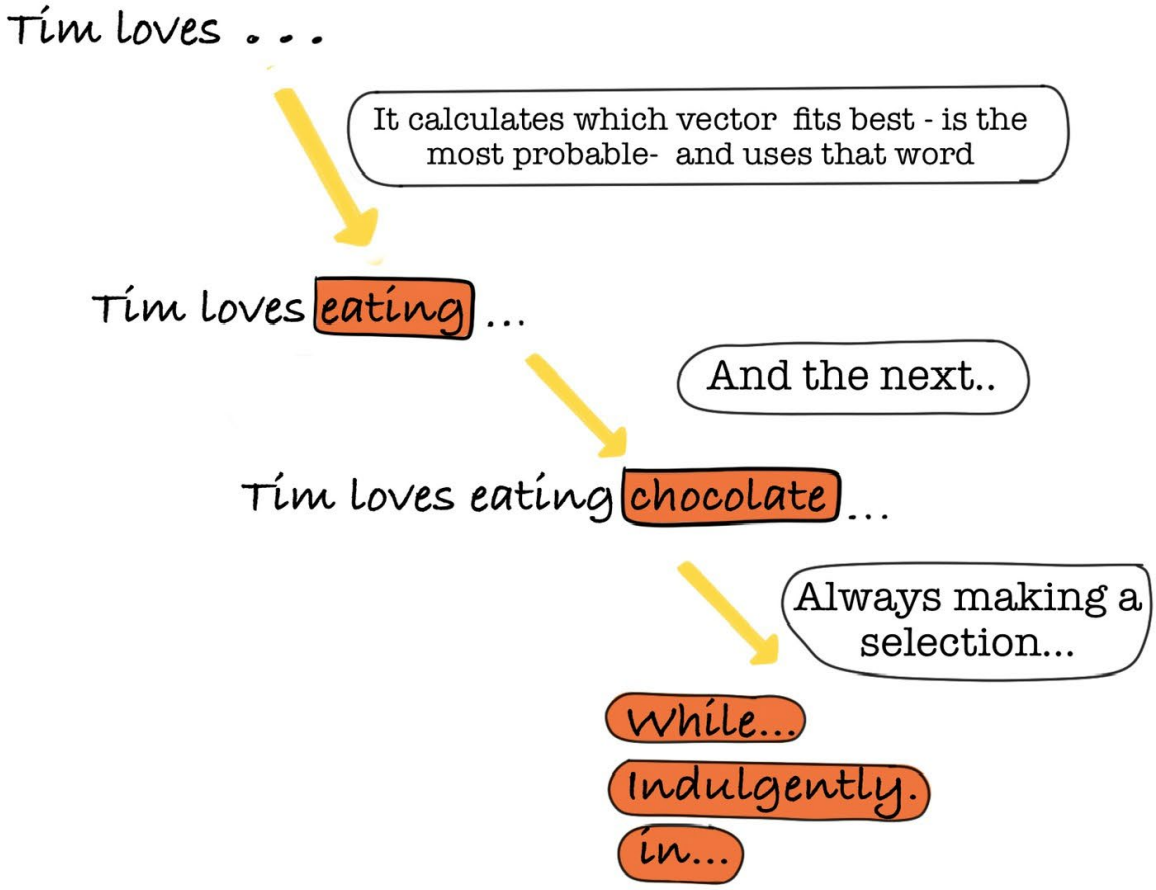
In large models, the vectors are made up of hundreds of numbers

These numbers together are called vectors. You can imagine it as arrows and the numbers give you the directions



Words with similar meaning have arrows (vectors) close to each other

When a chatbot answers you, it always predicts the next word.



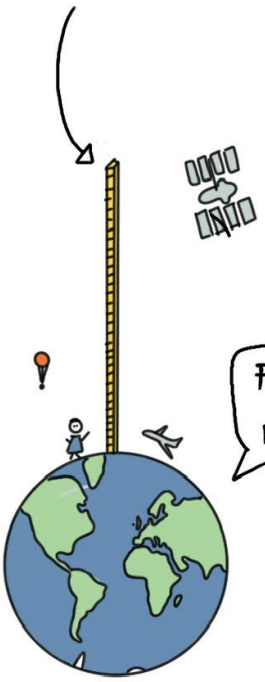
These models pay attention to multiple features (meaning, position in a sentence, style, etc) to create meaningful text, but

Don't forget! AI does not have an understanding of the words and the world as we do.

by Orsi Nagy

To learn languages, generative AI models are trained on billion words.

That is like 3,000,000 books



The AI uses all this data to build the model. This information is stored in parameters. It is then used to predict the next word.

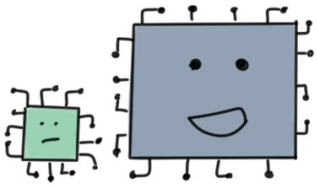
For each book you can read, LLMs have read 10.000!

But it matters WHAT you read too!

Is it good quality?

Does it bring knowledge from diverse cultures, languages, viewpoints?

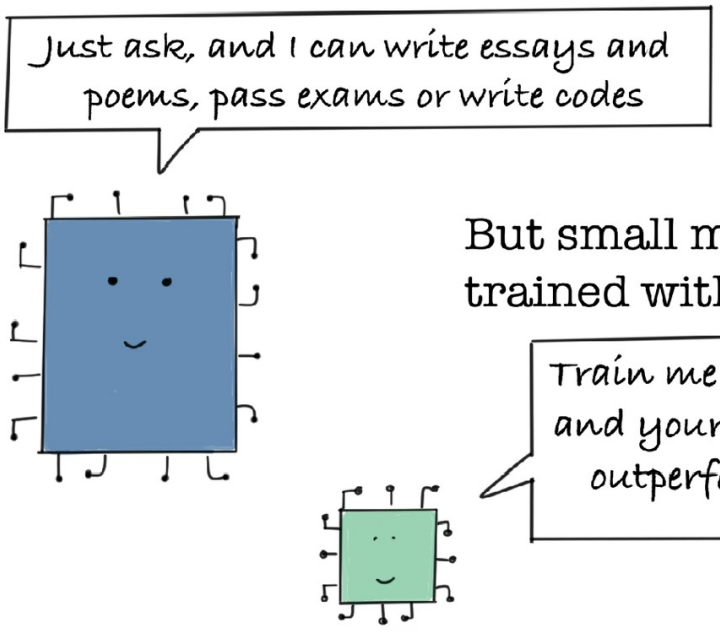
Bigger is better?



Large Language Models are large. We measure their size in parameters.

Numerical, statistical values that define how the text tends to flow
Chat GPT4: largest 1.7 trillion ~;
But Bloomberg's model works well with 50 billion parameters too

Big models can give you good answers across many fields.

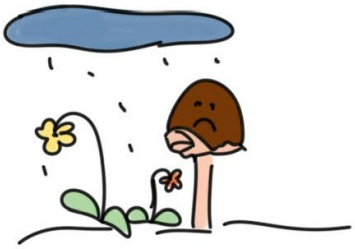


The downsize of a big model is that it costs a lot of money and energy to train and run them

Training even a smaller model can take as much energy as a transatlantic flight

One query takes as much energy as boiling two spoons of water

If a lot of people start using the model, we should be conscious of their environmental impact - very likely the pollution will be invisible for most users!



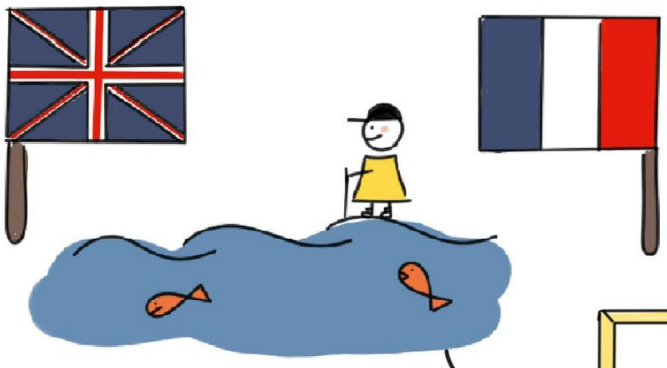
But always remember, the answer you get is made up of the most probable words.

It may be not even be true!

When an AI gives you answers that are not true, or made up, it's called hallucination.

Example by ChatGPT

Chris Wandelratsch walked for 14 hours 51 minutes on the sea from France to England.

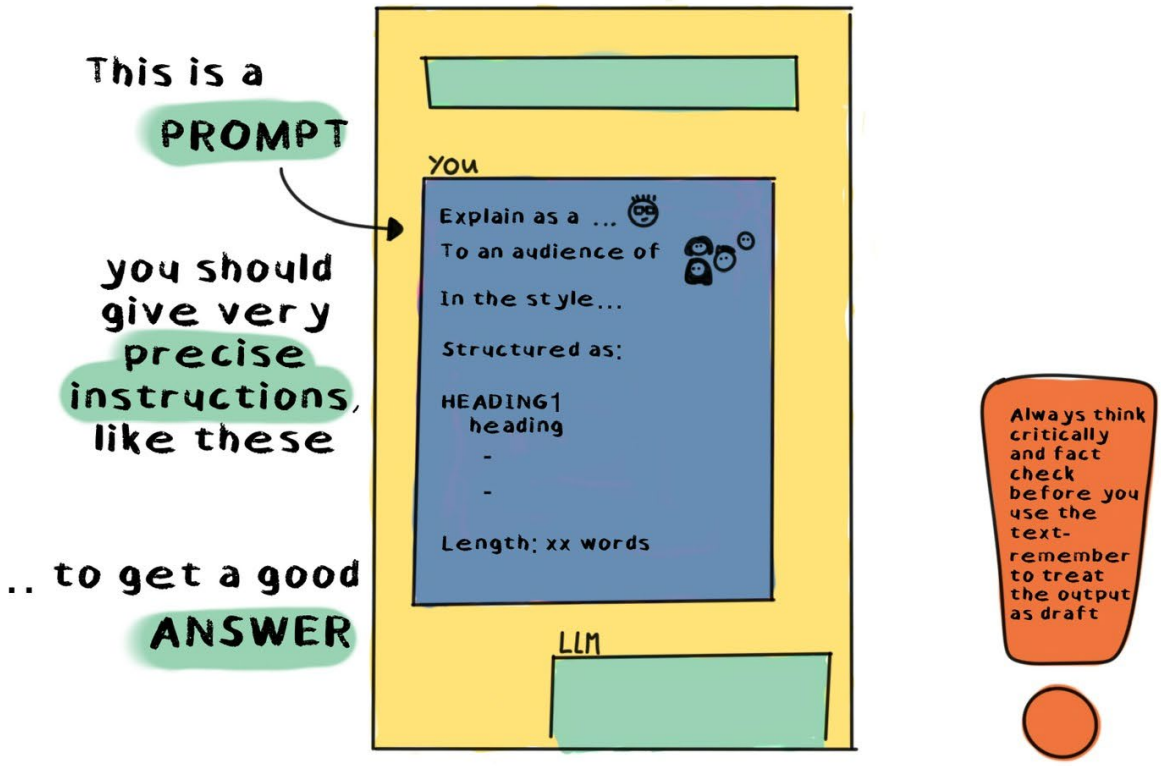


Thanks to recent improvements, such crazy hallucinations are rare now, but you can get misleading and false answers!

by Orsi Nagy

Prompt Engineering

techniques and tricks for writing your question so that you get more accurate answers



You can also use these strategies

Think of the exercise as chat. If the first answer is not perfect, you can go back and forth to ask for changes

For complex tasks, break down the task in steps. 'First, list all the actions for.... Second, highlight the actions without deadlines....'

-Ask for alternatives 'write 5 ideas for a post on...'

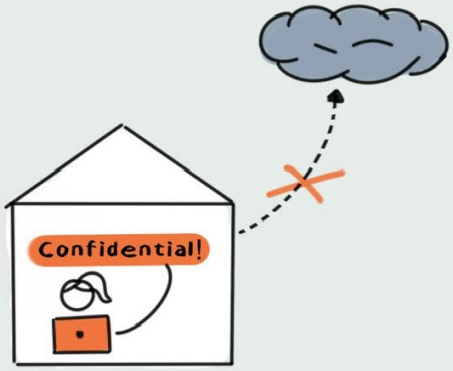
- give examples (few-shot prompting) 'take these two emails as example. Using their style, write an email to my boss ...'

- reverse: 'ask me questions until you have all info to write a cover letter for...'

Inspired by: Promt Engineering Specialisation on Coursera by Vanderbilt University

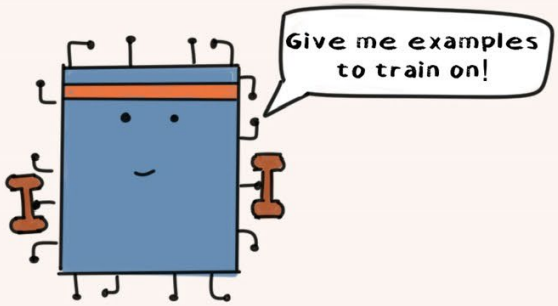
by Orsi Nagy

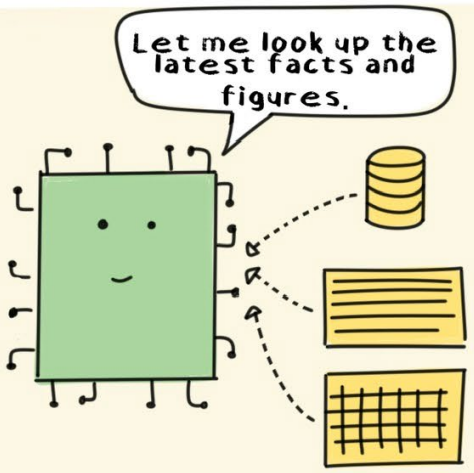
There are a number of technical improvements we can do so that LLMs give us better, more accurate answers:



Instead of working with an LLM on the cloud, we can host it in-house. This way we can use it with our own data without sharing it.

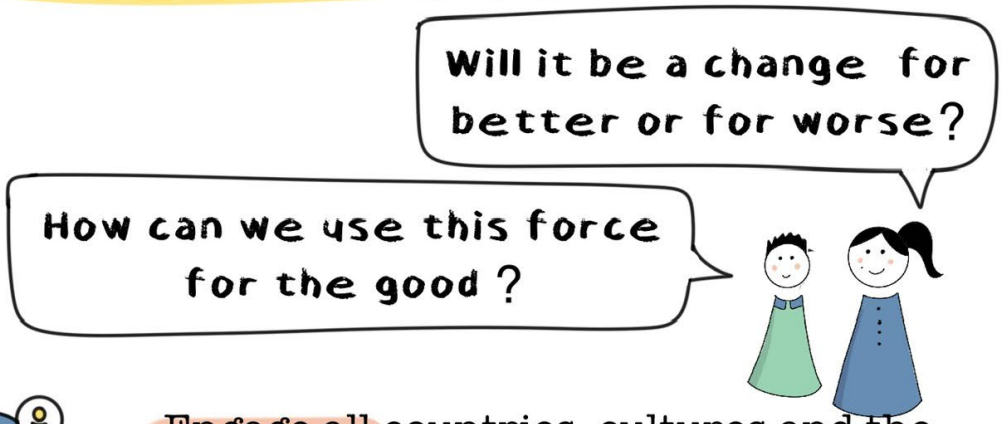
We can further train -fine-tune - a model on specific tasks or on a domain.



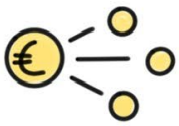


We can connect it to databases. This way we get responses that we know for a fact are true - and still get the fancy writing. (Retrieval Augmented Generation)

Large language models and AI is changing our world and will continue to do so.



Engage all countries, cultures and the entire society to define values, principles, and security concerns



Benefits for all: share economic and financial gains (e.g. benefit from reduced working hours)

Informed access to these systems: empower people to use these tools to improve our personal and professional lives

And remember: if you see a convincing text on a fancy page, it may have been written by humans or by machines

In any case, it's not necessarily TRUE!

by Orsi Nagy