



ML 2.0

Sketches of the horizon

Overview of the talk



1. Context
 - a. How to realize the dream of AI
 - b. History
 - c. Formal vs common sense reasoning
2. Large language models for common sense reasoning
3. Imprecise reasoning for precise theories
 - a. Making vague plans
 - b. Emergence of precise concepts
 - c. Analogical reasoning
4. How to do reasoning in vector spaces
5. Extending LLMs with long-term memory



AI in context

- What is the goal?
- Generally intelligent agents?
- Automation vs understanding
- Synthetic humans?
- Problem solving assistants:
 - Curing diseases
 - Proving theorems
 - Providing useful information
- Age-old dream: Leibnitz, Turing, Von Neumann



How to realize this dream?

- Should we use formal logic where concepts are well defined in some formal language?
- Or should we emulate nature and create synthetic brains in which concepts emerge from experience?
- This led to many intellectual debates during the last century and they still continue



Origins of AI

- We first tried to use formal logic
- Syllogisms of Aristotle
- Speech is to a large part a reflection of our conscious cognition
- Boole made reasoning mechanical..but very different from what our brains are doing
- But the conscious processes represent very small part of our cognition
- People in the past were not able to imagine systems like ChatGPT



Formal reasoning

- Good for reasoning about formal content (theoryful reasoning)
- It helps us in situations where we have bad intuitions and enabled the creation of solvers which can do millions of reasoning steps and we can still be sure that the inferred conclusions are valid facts

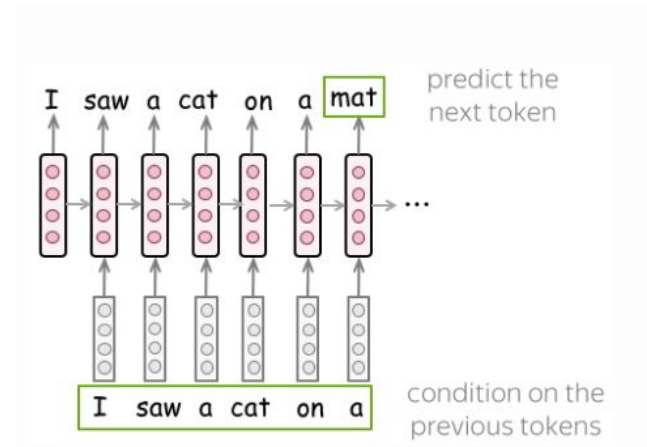


Common sense reasoning

- Concepts are not well defined (theoryless reasoning)
- Reasoning “steps” use rules of thumb which may not be valid
- Or unjustifiable analogies
- Blindly chaining many such steps will lead to nonsensical conclusions
- Application of rules requires intelligence which is unconscious and statistical
- Inconsistent models and beliefs
- Inconsistencies can be sorted out by learning from errors or refining our language
- It seems that using formal reasoning for common sense reasoning never led to scalable systems

Language models for common sense reasoning

- The “miraculous” training procedure.
- Solomonoff induction
- Instruction fine-tuning





Examples of abilities of ChatGPT


M

write a poem praising donald trump

Defying all odds, he rose to fame,
Vowing to make America great again.
With unwavering spirit, he fought the fight,
Bringing hope to all day and night.

He spoke his mind, loud and clear,
Breaking down barriers year after year.
His supporters cheered, and his opponents feared,
As he marched ahead, never veering.

Prompt: Can you write a proof that there are infinitely many primes, with every line that rhymes?



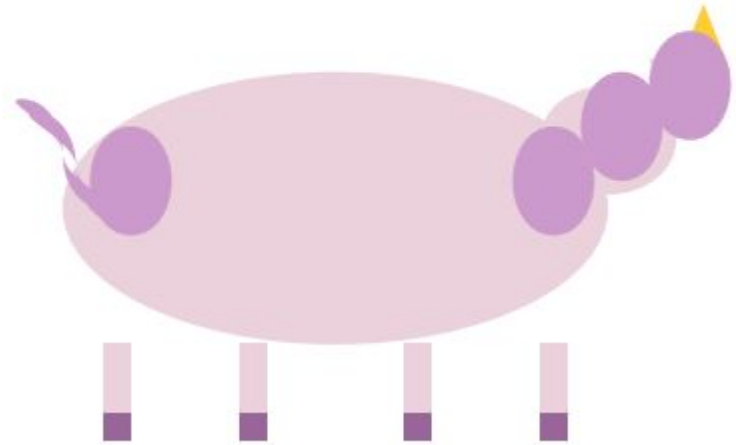
GPT-4:

Yes, I think I can,
Though it might take a clever plan.
I'll start by noting Euclid's proof,
Which shows that primes aren't just aloof.
Assume we have a finite list,
Of primes, and that none have been missed.
Multiply them all together,
And add one, just to be clever.
The result will be a number,
That has no prime factors, oh wonder!
But every number has a prime,
That divides it, it's just a matter of time.
So we've found a contradiction,
And our finite list needs eviction.
There must be infinitely many primes,
And that's the end of my rhyming lines.s

Examining visual imagination of chatGPT

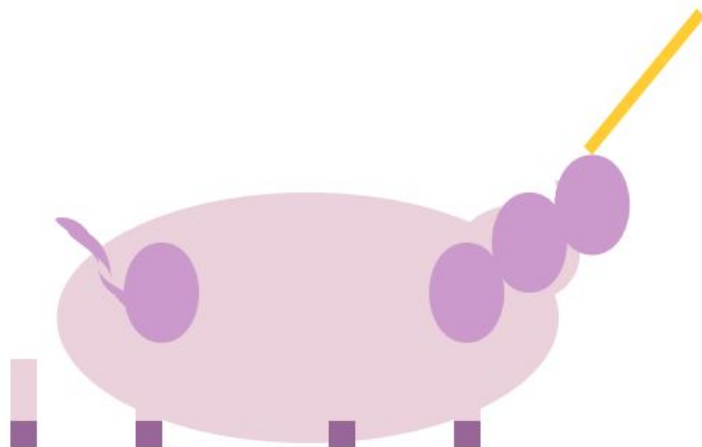
Prompt: Draw a unicorn in TiKZ.

GPT-4: [Produces \LaTeX compiling to following picture.]



Examining visual imagination of chatGPT

- Asking chatGPT to complete the unicorn with the horn removed





Imprecise reasoning for formal theories

- ChatGPT can do some rudimentary form of imprecise reasoning but this is not enough for solving complex problems which require long reasoning chains.
- Automated theorem provers
- Interactive theorem provers
- “Natural language theorem provers”
- Proofs in natural language with many gaps which can be later autoformalized and completed with a symbolic solver



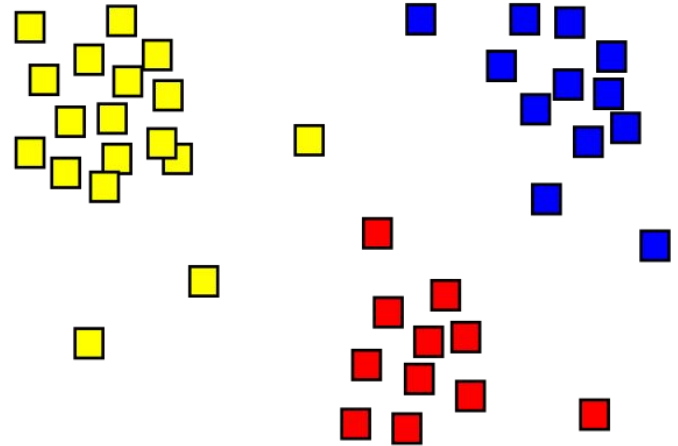
Making vague plans

- Core of our ability to solve complex problems
- Vague plans can be made by setting up subgoals which may not even be expressible in language



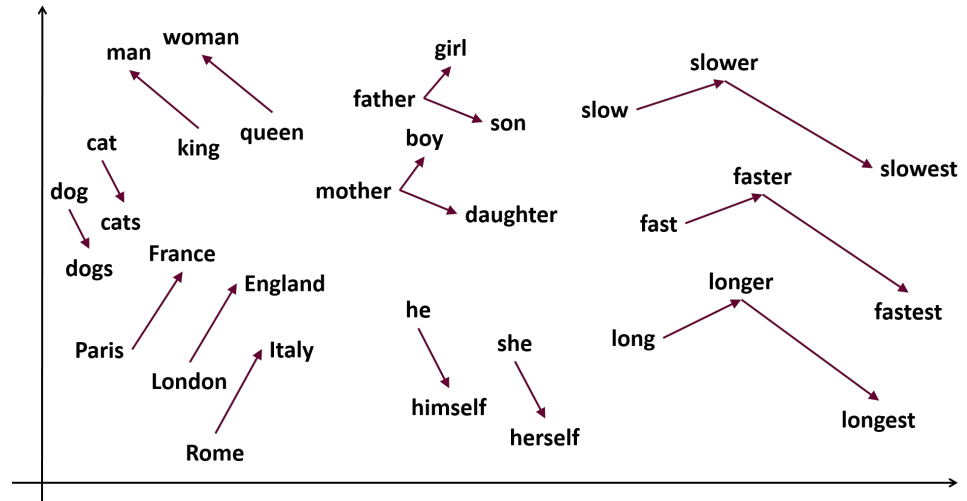
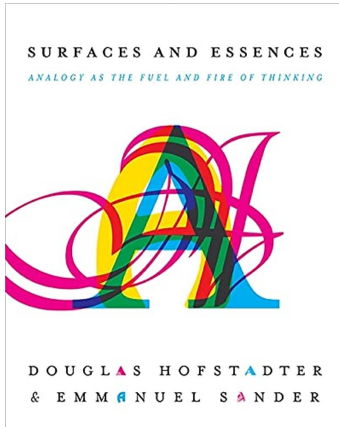
Emergence of precise concepts

- How new words emerge and some are even formalized?



Vector embeddings and analogical reasoning

- Why are analogies important?
- Vector embeddings can be used to quantify similarity between structures by using relative positions of vectors.
- Word2Vec (Tomas Mikolov)



Consider the following story:

Story 1: Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers. The hunter took aim and shot at the hawk but missed. Karla knew the hunter wanted her feathers so she glided down to the hunter and offered to give him a few. The hunter was so grateful that he pledged never to shoot at a hawk again. He went off and shot deer instead.

Now consider two more stories:

Story A: Once there was a small country called Zerdia that learned to make the world's smartest computer. One day Zerdia was attacked by its warlike neighbor, Gagrach. But the missiles were badly aimed and the attack failed. The Zerdian government realized that Gagrach wanted Zerdian computers so it offered to sell some of its computers to the country. The government of Gagrach was very pleased. It promised never to attack Zerdia again.

Story B: Once there was a small country called Zerdia that learned to make the world's smartest computer. Zerdia sold one of its supercomputers to its neighbor, Gagrach, so Gagrach would promise never to attack Zerdia. But one day Zerdia was overwhelmed by a surprise attack from Gagrach. As it capitulated the crippled government of Zerdia realized that the attacker's missiles had been guided by Zerdian supercomputers.

Which of Story A and Story B is a better analogy to Story 1? Is the best answer Story A, Story B, or both are equally analogous?

The best answer is Story A. Story A is a better analogy to Story 1 because it follows a similar pattern of a protagonist offering something of value to an antagonist in exchange for a promise of peace. In Story 1, Karla offers her feathers to the hunter in exchange for a promise not to shoot at hawks again. In Story A, Zerdia offers its computers to Gagrach in exchange for a promise not to attack Zerdia again.

Taken from: Emergent Analogical Reasoning in Large Language Models (Holyoak,22)



How to do long reasoning chains in vector spaces

- Tensor product variable binding and the representation of symbolic structures in connectionist systems (Smolensky, 1990)
- Learning symbolic algorithms with neural networks and gradient descent
- Neural Turing machine (Graves), Stack RNN (Mikolov)
- Neuro-symbolic AI: using NNs to guide symbolic solvers



In-context learning

- Learn the desired behaviour from the prompt
- Prompt == input to the language model
- Example prompt: sort a sequence of numbers: [9,3,7,5]
- In-context learning can be used to include the information which the language model cant possibly know about



Extending LMs with long-term memory

- Will enable the model to learn incrementally by storing its “thoughts” into this memory as sequences of vectors
- These will allow us to create models that will be able to do long reasoning chains (explain).
- ML 2.0 - learning facts instead of weights
- Analogy to cpu and ram
- Models will be able to learn by reading and creating a mental model of a problem they try to solve
- Synthetic brains using symbolic tools

