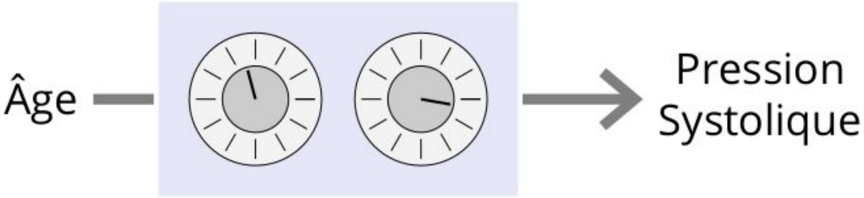
From deep learning to AI

NRB AI & Data Xperience, October 15

Prof. Damien Ernst, Prof. Gilles Louppe, Lize Pirenne

LIÈGE université

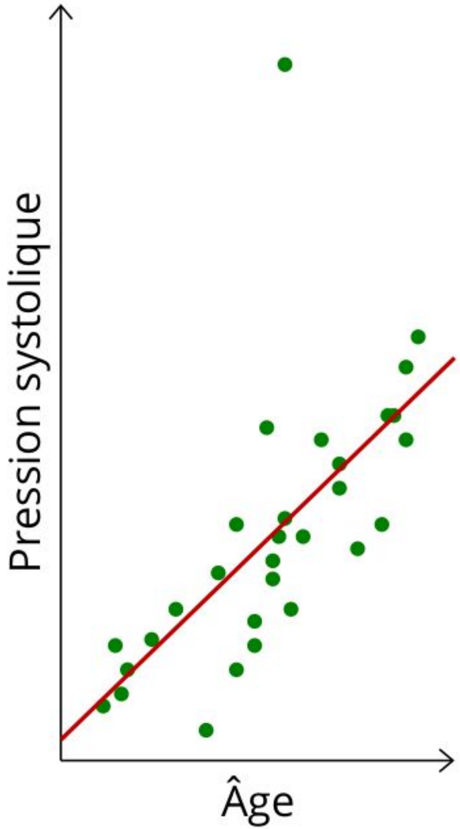
“For the last forty years we have programmed computers; for the next forty years we will train them.” Christopher Bishop, 2020.



Âge

Pression Systolique

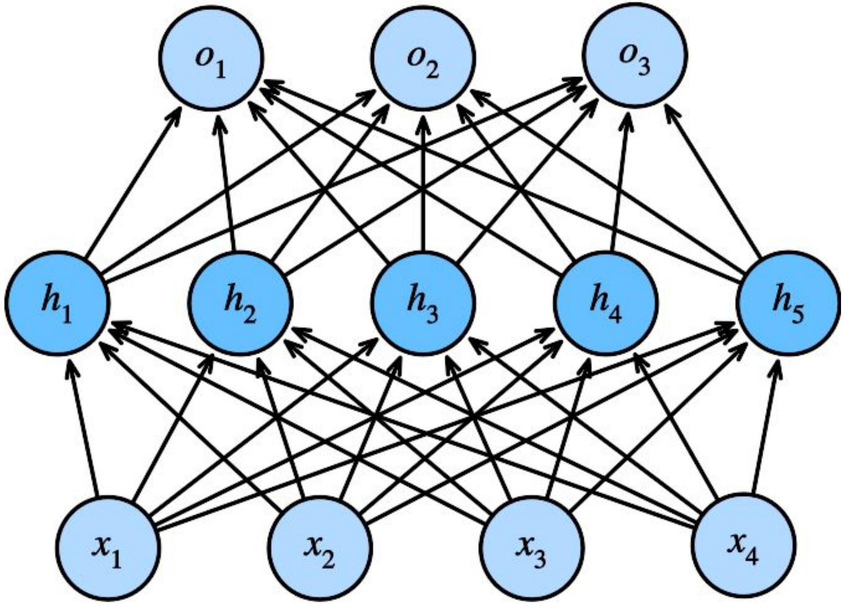
“Régression linéaire”



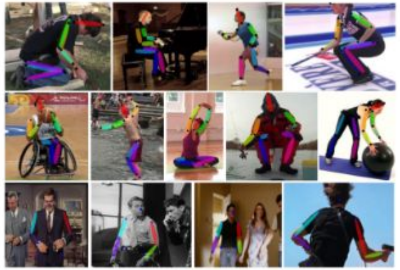
The **machine learning** approach to problem-solving.

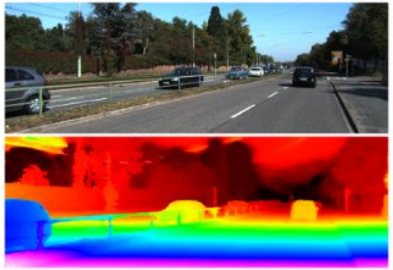
Deep learning scales up the machine learning approach by

- using larger models known as **neural networks**,
- training on **larger datasets**,
- using **more compute** resources.



Specialized neural networks can be trained to achieve **super-human performance** on many complex tasks that were previously thought to be out of reach for machines.



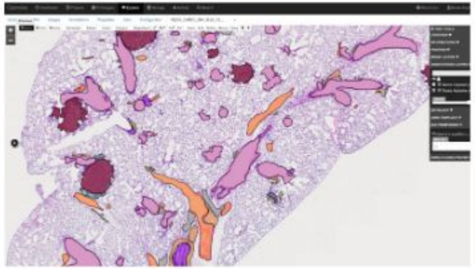


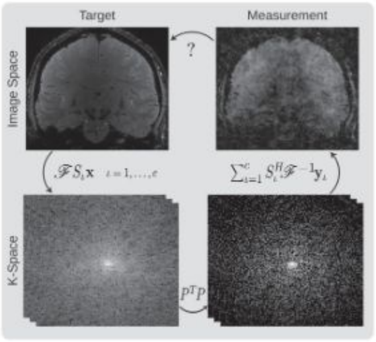
A person riding a motorcycle on a dirt road.
Two dogs play in the grass.
A group of young people playing a game of frisbee.
Two hockey players are fighting over the puck.

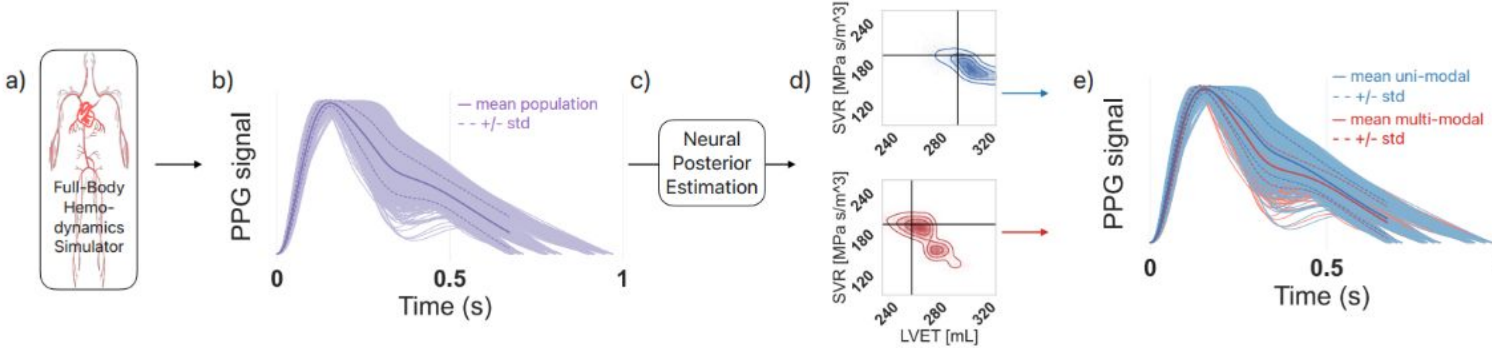
I: Jane went to the hallway.
I: Mary walked to the bathroom.
I: Sandra went to the garden.
I: Daniel went back to the garden.
I: Sandra took the milk there.
Q: Where is the milk?
A: garden

(Top) Scene understanding, pose estimation, geometric reasoning.
(Bottom) Planning, Image captioning, Question answering.

Neural networks form **primitives** that can be transferred to many domains.

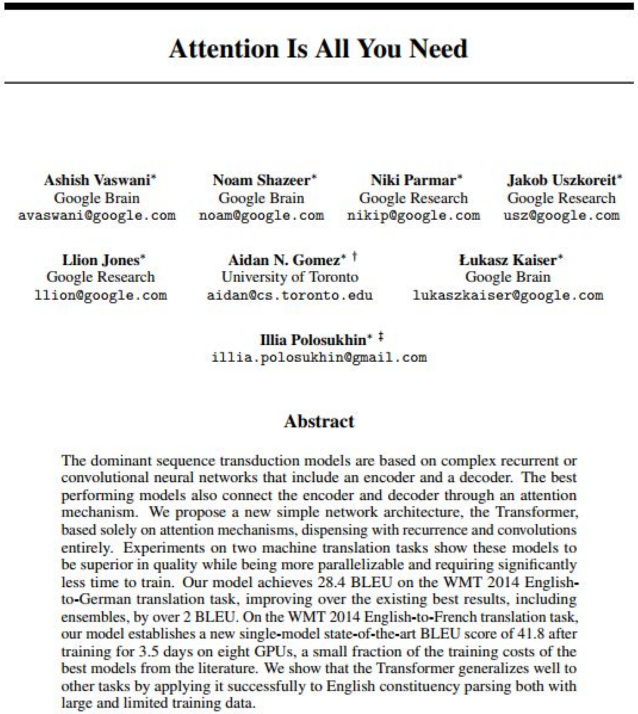



Target
Measurement
Image Space
?
$\mathscr{F} S_i \mathbf{x} \quad i = 1, \ldots, c$
$\sum_{i=1}^{c} S_i^H \mathscr{F}^{-1} \mathbf{y}_i$
K-Space
$P^T P$



(Top) Analysis of histological slides, denoising of MRI images, nevus detection.
(Bottom) Whole-body hemodynamics reconstruction from PPG signals.

The breakthrough



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

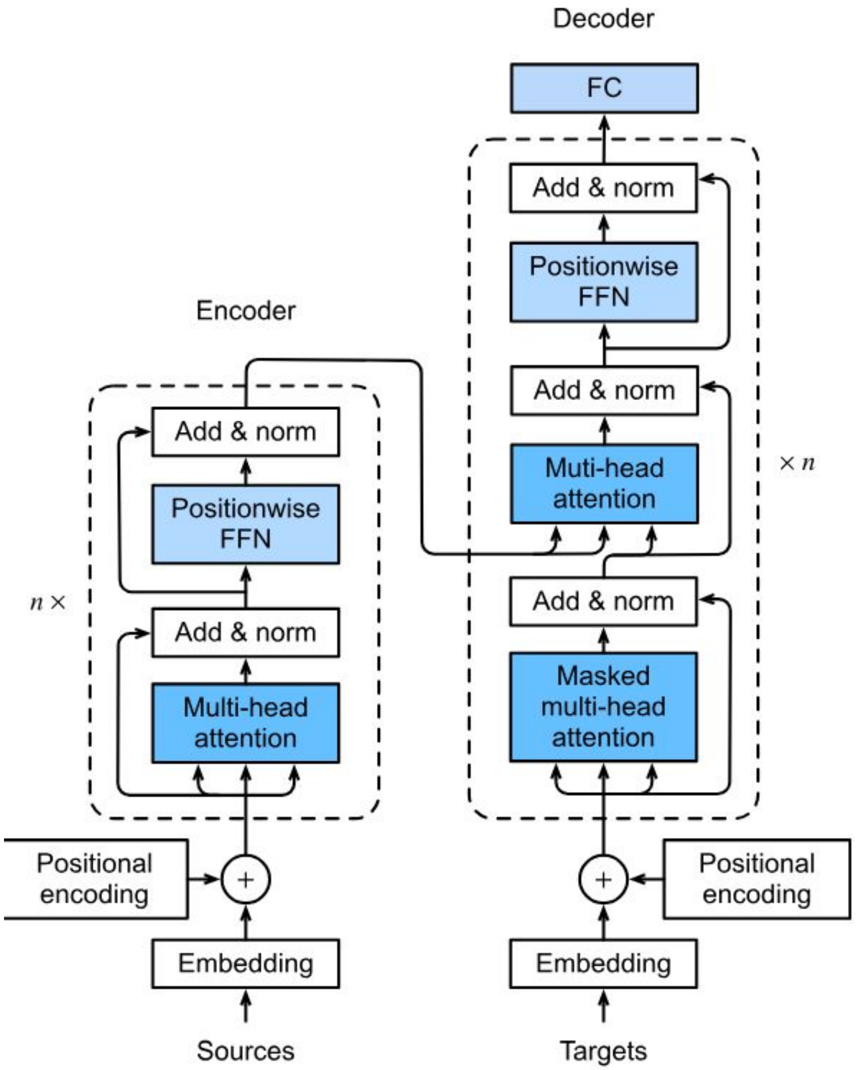
Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

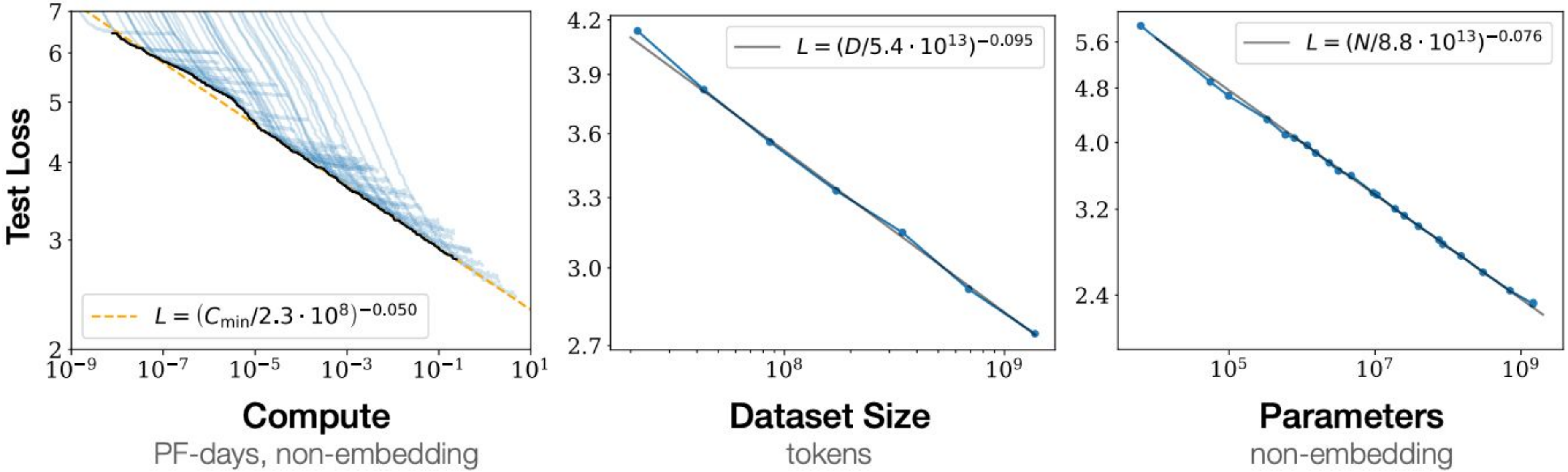
Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Vaswani et al, 2017.





A **brutal simplicity**:

- The more data, the better the model.
- The more parameters, the better the model.
- The more compute, the better the model.

Conversational assistants (Anthropic, 2024)

Code assistants (Cursor, 2024)

Multi-modal assistants (OpenAI, 2024)

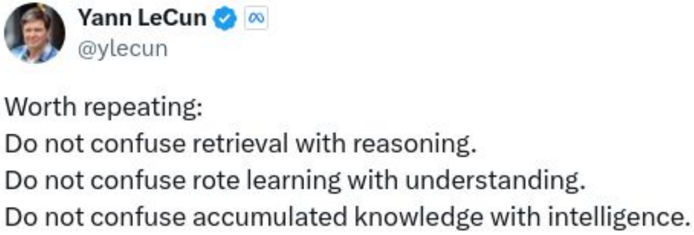
New opportunities

If Bob knew all of human knowledge and understood* everything you do, say or write…

- What questions would you ask Bob?
- What would do with Bob if you could hire him in your company?
- What if all your employees could talk with Bob?
- What if all your customers could talk with Bob?
- What if Bob was also working for all your competitors?

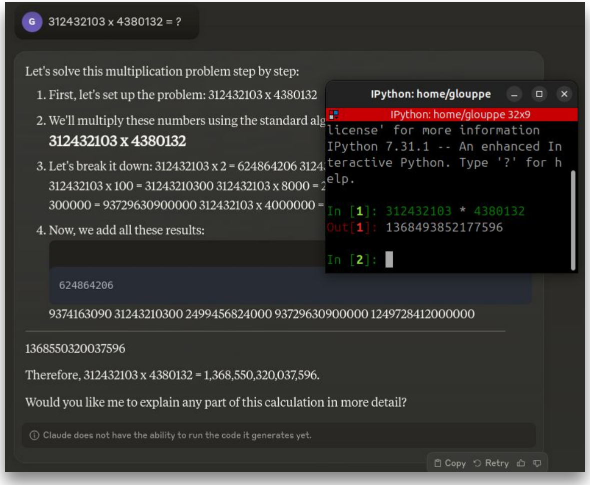
Limitations

- LLMs give the impression of understanding.
- LLMs show no evidence of formal reasoning.
- LLMs are fragile. Rephrasing questions alter results significantly.
- LLMs do not know what do they not know.



Yann LeCun
@ylecun

Worth repeating:
Do not confuse retrieval with reasoning.
Do not confuse rote learning with understanding.
Do not confuse accumulated knowledge with intelligence.



312432103 x 4380132 = ?

Let's solve this multiplication problem step by step:

1. First, let's set up the problem: 312432103 x 4380132
2. We'll multiply these numbers using the standard alg
   312432103 x 4380132
3. Let's break it down: 312432103 x 2 = 624864206 3124
   312432103 x 100 = 31243210300 312432103 x 8000 = 2
   300000 = 93729630900000 312432103 x 4000000 =
4. Now, we add all these results:

```
624864206
```

9374163090 31243210300 2499456824000 93729630900000 1249728412000000

1368550320037596

Therefore, 312432103 x 4380132 = 1,368,550,320,037,596.

Would you like me to explain any part of this calculation in more detail?

Claude does not have the ability to run the code it generates yet.

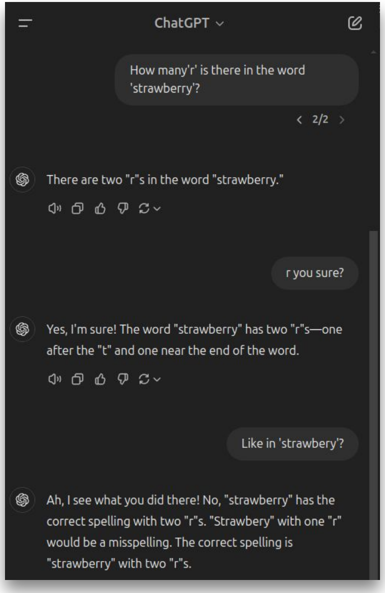
IPython: home/glouppe

IPython: home/glouppe 32x9

```
license' for more information
IPython 7.31.1 -- An enhanced In
teractive Python. Type '?' for h
elp.

In [1]: 312432103 * 4380132
Out[1]: 1368493852177596

In [2]: 
```



How many 'r' is there in the word 'strawberry'?

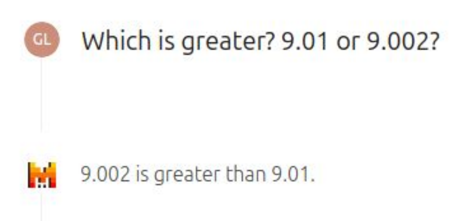
There are two "r"s in the word "strawberry."

r you sure?

Yes, I'm sure! The word "strawberry" has two "r"s—one after the "t" and one near the end of the word.

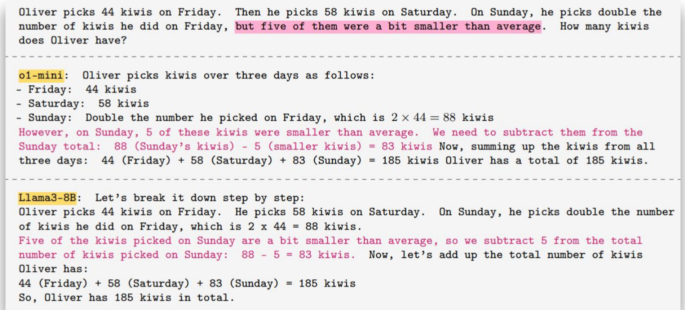
Like in 'strawbery'?

Ah, I see what you did there! No, "strawberry" has the correct spelling with two "r"s. "Strawbery" with one "r" would be a misspelling. The correct spelling is "strawberry" with two "r"s.



Which is greater? 9.01 or 9.002?

9.002 is greater than 9.01.

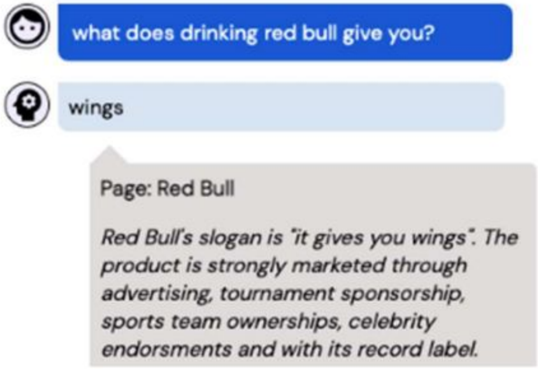


```
Oliver picks 44 kiwis on Friday.  Then he picks 58 kiwis on Saturday.  On Sunday, he picks double the
number of kiwis he did on Friday, but five of them were a bit smaller than average.  How many kiwis
does Oliver have?
```

```
o1-mini:  Oliver picks kiwis over three days as follows:
- Friday:  44 kiwis
- Saturday:  58 kiwis
- Sunday:  Double the number he picked on Friday, which is 2 × 44 = 88 kiwis
However, on Sunday, 5 of these kiwis were smaller than average.  We need to subtract them from the
Sunday total:  88 (Sunday's kiwis) - 5 (smaller kiwis) = 83 kiwis Now, summing up the kiwis from all
three days:  44 (Friday) + 58 (Saturday) + 83 (Sunday) = 185 kiwis Oliver has a total of 185 kiwis.
```

```
Llama3-8B:  Let's break it down step by step:
Oliver picks 44 kiwis on Friday.  He picks 58 kiwis on Saturday.  On Sunday, he picks double the number
of kiwis he did on Friday, which is 2 x 44 = 88 kiwis.
Five of the kiwis picked on Sunday are a bit smaller than average, so we subtract 5 from the total
number of kiwis picked on Sunday:  88 - 5 = 83 kiwis.  Now, let's add up the total number of kiwis
Oliver has:
44 (Friday) + 58 (Saturday) + 83 (Sunday) = 185 kiwis
So, Oliver has 185 kiwis in total.
```

LLMs can be made more robust and helpful if the **rationale** of their answers can be **extracted or explained**.



what does drinking red bull give you?

wings

Page: Red Bull

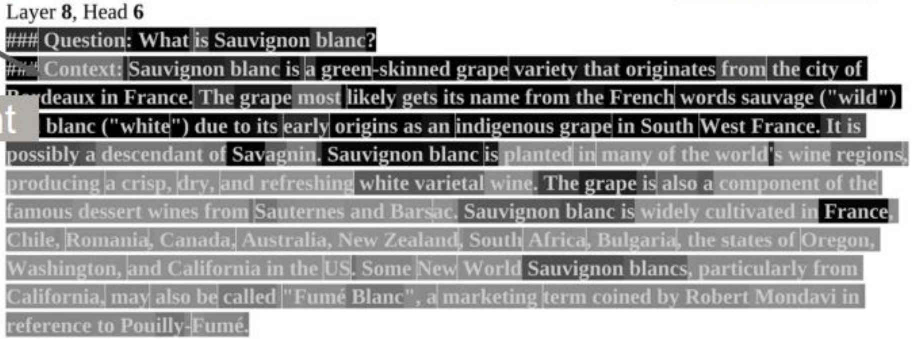
Red Bull's slogan is "it gives you wings". The product is strongly marketed through advertising, tournament sponsorship, sports team ownerships, celebrity endorsments and with its record label.

Pirenne et al, 2024.

Our solution was to use the inner workings of an LLM to see what it pays attention to. More specifically, we were analysing the attention matrix to find the sentences in the context needed for generating the answer.

Rationale In black

Irrelevant in grey

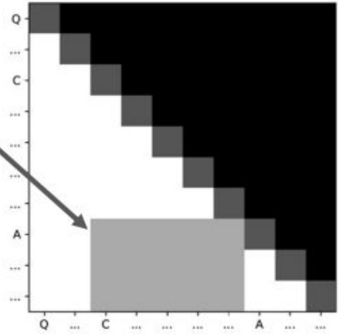


Layer **8**, Head **6**

Question: What is Sauvignon blanc?

Context: Sauvignon blanc is a green-skinned grape variety that originates from the city of Bordeaux in France. The grape most likely gets its name from the French words sauvage ("wild") blanc ("white") due to its early origins as an indigenous grape in South West France. It is possibly a descendant of Savagnin. Sauvignon blanc is planted in many of the world's wine regions, producing a crisp, dry, and refreshing white varietal wine. The grape is also a component of the famous dessert wines from Sauternes and Barsac. Sauvignon blanc is widely cultivated in France, Chile, Romania, Canada, Australia, New Zealand, South Africa, Bulgaria, the states of Oregon, Washington, and California in the US. Some New World Sauvignon blancs, particularly from California, may also be called "Fumé Blanc", a marketing term coined by Robert Mondavi in reference to Pouilly-Fumé.

~ Auto-correlation of the input



The research challenges of deploying LLMs for sensitive applications (e.g., medical, legal, financial) highlight **opportunities for mutually beneficial collaborations between the industry and academic**. (Many others exist, beyond LLMs!)

Benefits **for academia**:

- Access to relevant real-world problems.
- Access to industry-grade computing resources for research.
- Access to real-world data.
- Financial support.

Benefits **for industry**:

- Solutions for problems requiring extensive research efforts.
- Direct access to up-to-date knowledge in AI.
- Opportunities to work with brilliant minds and potential future employees.

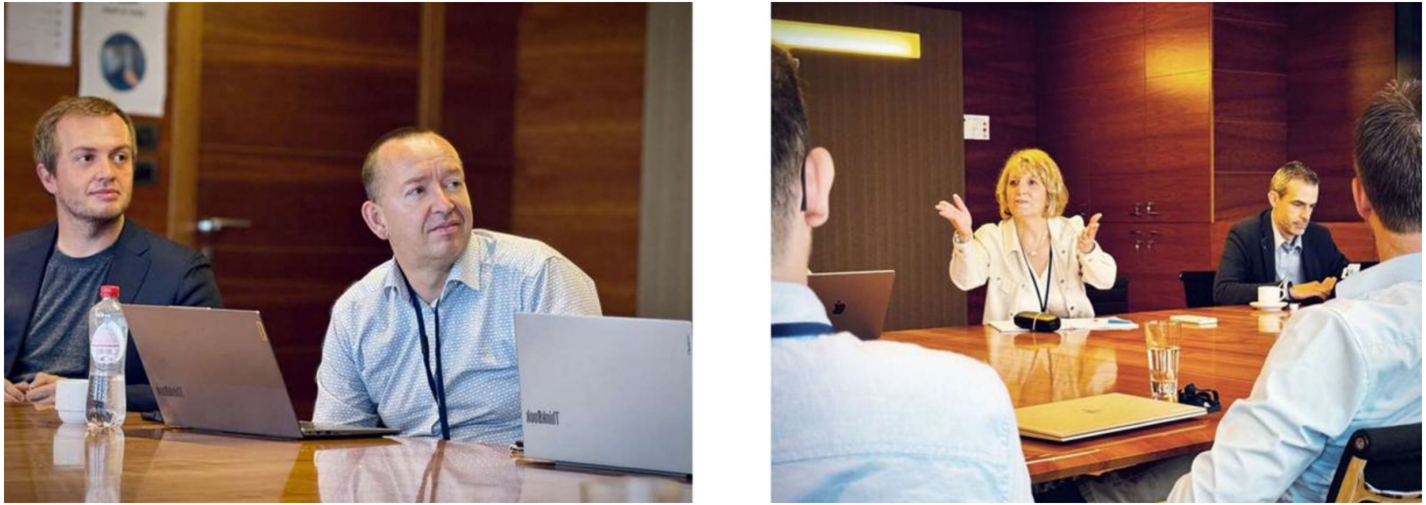
NRB - ULiège Research Chair

LIÈGE université

NRB

DARING TO COMMIT

- 4-year funding for research on **AI for the industry**.
- Prof. Damien Ernst (LLMs, Reinforcement Learning, Energy)
  Prof. Gilles Louppe (Deep Learning, Generative AI, Digital Twins, AI4Science)
- Joint teams, regular meetings and shared objectives to encourage synergies.



(Left) Prof. Louppe and Prof. Ernst, (Right) Laurence Mathieu, CEO of NRB.

AI & Data Xperience