



# Introduzione al Machine Learning (...altro che intelligenza)

**Giuseppe Lo Presti**  
*CERN IT Department*

*Italian Teachers Programme 2024 - Discovery*



Start a post, try writing with AI



Media



Contribute expertise



Write article



Mathesis  
Società italiana di scienze  
matematiche e fisiche  
fondata nel 1895

La MATEMATICA

al TEMPO

dell'INTELLIGENZA ARTIFICIALE

L'impatto dell'IA sui processi educativi e formativi

Compresso  
MATHESIS

Roma 2023

## Pricing

\$5 / 100 credits

Average Cost and Time

Basic Deepfake (10,000 Iterations) - \$15

High-Quality Deepfake (50,000 Iterations) - \$60

Create a Deepfake Video

\* The quality of your deepfake depends on the quality and length of your videos. The more well trained your AI model is, the better the result.  
\* Please use our technology responsibly and review our [content policy](#) and our [terms of use](#).



Copilot, 22:48

You're now approved! You can use Bing in Skype, Bing.com, Edge, or Bing Mobile.

Ask me to check weather, write you a poem, or find that movie title you forgot. My responses are AI-generated, so be sure to check the facts.

FIU Business

About Us Hire our Students Contact Us Give [Apply now](#)

Academics Students Alumni Career Services Faculty & Research News & Events Giving

Overview Programs Directory International Students Financial Aid More

School Insights

ice

# Artificial intelligence in the stock market: how did it happen?

Opinion Sport Culture Lifestyle

The Guardian

Site crisis Ukraine Environment Science Global development Football Tech Business Obituaries

Intelligence

## AI's 'Oppenheimer moment': autonomous weapons enter the battlefield



Scientific Background to the Nobel Prize in Physics 2024

“FOR FOUNDATIONAL DISCOVERIES AND INVENTIONS THAT ENABLE MACHINE LEARNING WITH ARTIFICIAL NEURAL NETWORKS”

The Nobel Committee for Physics



G. Lo Presti - Italian Teachers

Nick Robins-Early

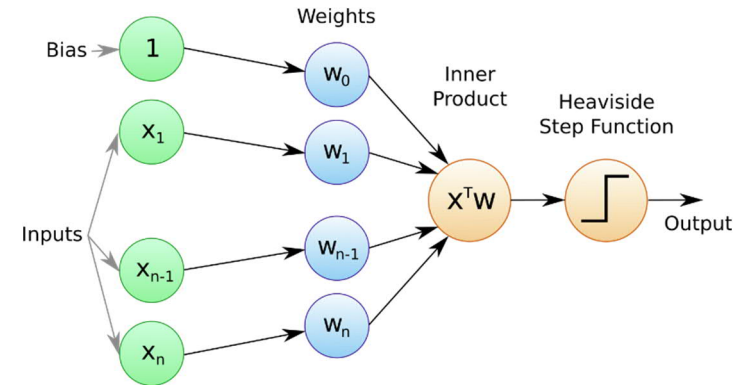
Sun 14 Jul 2024 18:00 CEST

# Big Data: a definition

- *Big Data* is a field that treats of ways to analyse [...] or otherwise deal with data sets that are **too large or complex to be dealt with** by traditional data-processing application software (*Wikipedia*)
  - **Moving target** by definition!
- From **structured** data, relational DBs, centralized processing...
- To **unstructured** data and decentralized (i.e. parallel and loosely-coupled) processing, more adapted to the Cloud
  - E.g. **trend analysis**, **pattern recognition**, **image segmentation**, **natural language interpretation/translation (ChatGPT!)**, ...

# The Power of Data

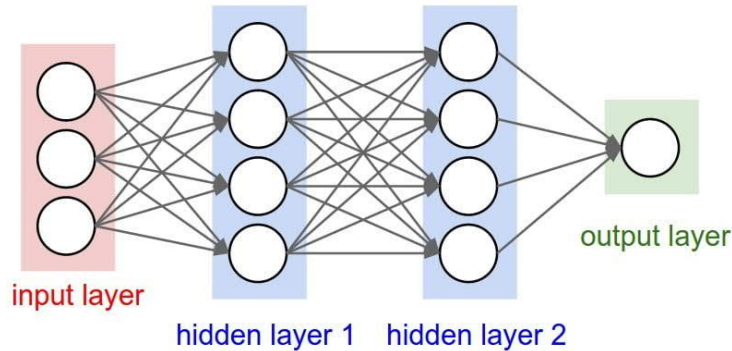
- **Neural Networks** are well known since the 1960s, but it's only now with **very large** and **easily accessible** data sets that they become effective!
- They are all based on simple “units”, such as the **perceptron** [Rosenblatt, 1958]
  - The weights  $w_i$  can be iteratively estimated (the **learning** phase) by imposing the outputs for several given inputs (*backpropagation*)
  - We may also have **unsupervised learning**, where the learning phase is partly automated



$$y = S(w_0 + \sum_i x_i w_i)$$

# Diving Deeper: Machine Learning

- Perceptrons may be connected in multiple layers



- Software frameworks are readily available to implement many configurations for **Deep Machine Learning**



PYTORCH

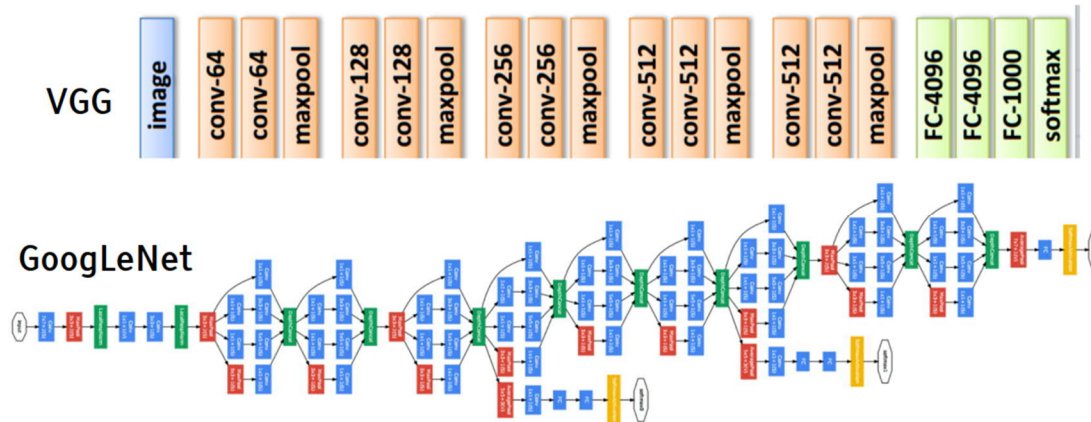
Deep Learning with PyTorch





# How Deep?

- Example: image tagging via *convolutional* networks
  - Thousands of layers, **millions** of parameters
  - Facebook: a billion pictures per day goes through such networks, which delivers its result within ~2 seconds



# How Deep?

- Example: natural language generation
  - Use of **Generative Pre-trained Transformers** to speed up the training phase
    - Transformers were proposed by [Google in 2017](#)
    - Heavily used since then: *demonstrator* [recently published](#) for training purposes
    - 2023: ChatGPT-4o estimated at a **trillion** parameters!
- **Large Language Models (LLMs)** for encapsulating domain-specific knowledge
  - Being prototyped at CERN-IT to help Support and Service Desk

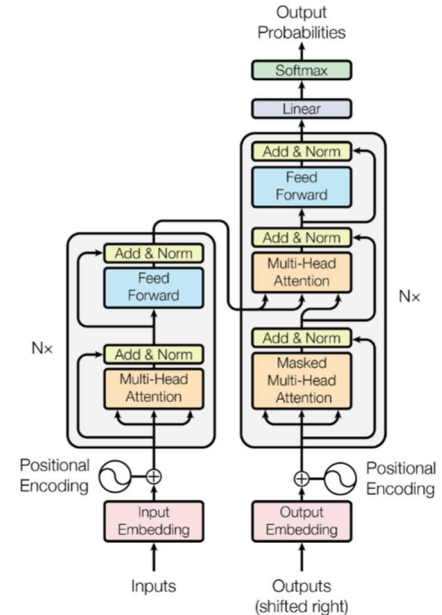
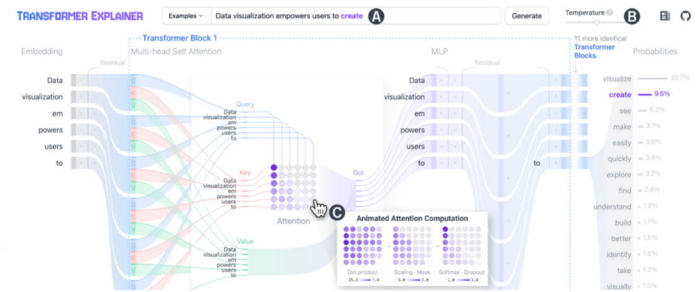


Figure 1: The Transformer - model architecture.



# New frontiers: Heterogeneous Computing

- (Deep) Machine Learning is so **crucial** that industry has long invested into **hardware acceleration**
  - **GPUs** (Graphical Processing Units), designed for videogames, now heavily used for faster matrix computations
  - **TPUs** (Tensor Processing Units), developed by Google, are offered in the Google Cloud Platform

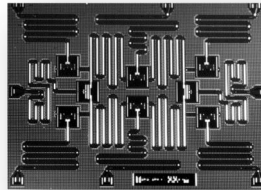




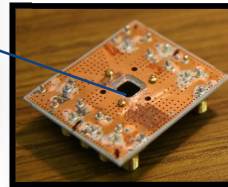
# New frontiers: Heterogeneous Computing

- A potential game changer: **Quantum Computing**
  - Quantum Computers can only execute a **very limited set of “programs”**, but with **exponential parallelism** (on paper)
  - **Quantum Machine Learning** is being demonstrated – also at CERN – as one of those programs, which can be executed by such hardware

Qubits on chip



Circuit board



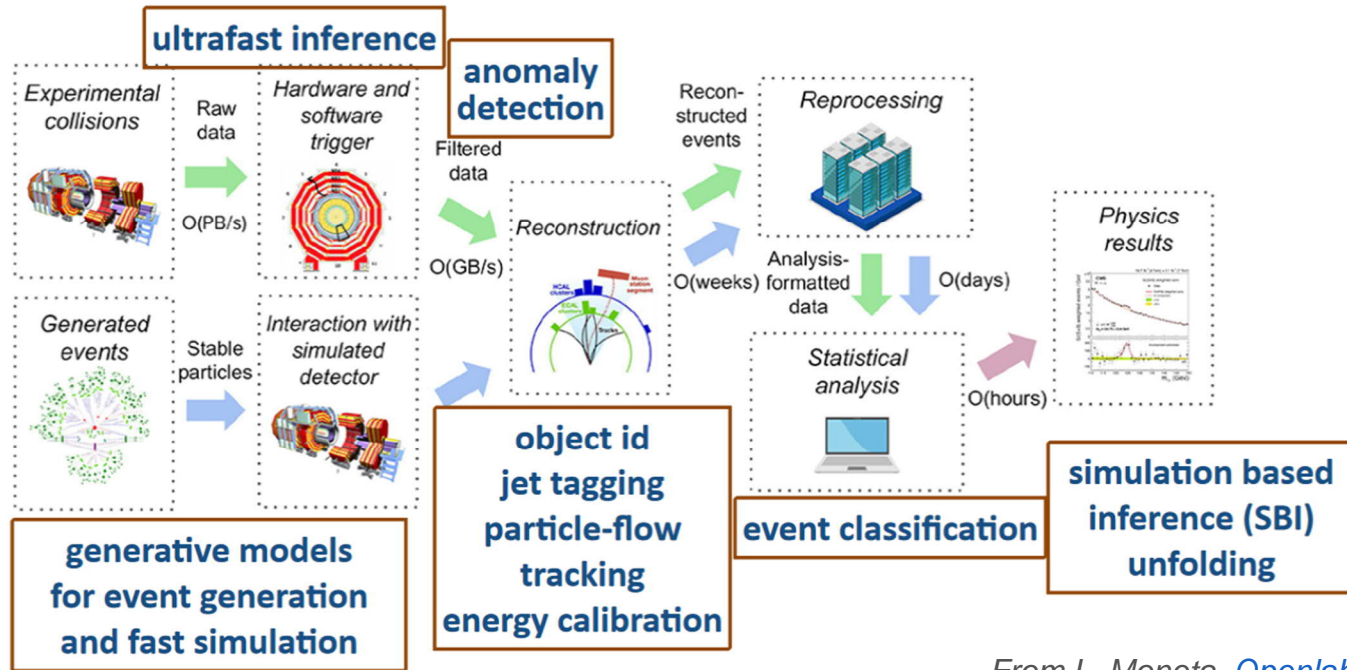
15 mK



Courtesy M. Grossi

# Machine Learning for Particle Physics

Inter-experiment ML working group to coordinate such activities



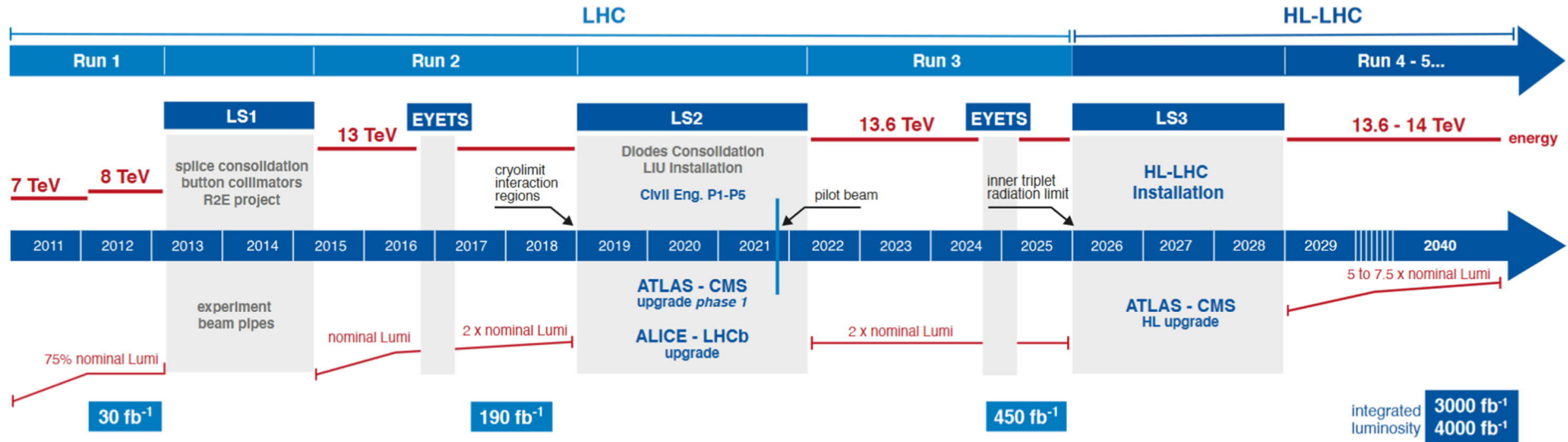
From L. Moneta, [Openlab workshop 2024](#)

# Machine Learning at CERN and beyond

- ML applied to **extract trends, detect or predict failures, detect anomalies (new Physics?), ...**
  - Control Systems: LHC Beams Control Logging
  - Security forensics, system analysis/profiling, etc.
  - Astronomy and other Big Sciences
- In general, ML techniques implemented where analytical approaches are **inapplicable/unpractical**



# Hi-Lumi LHC: a computing challenge



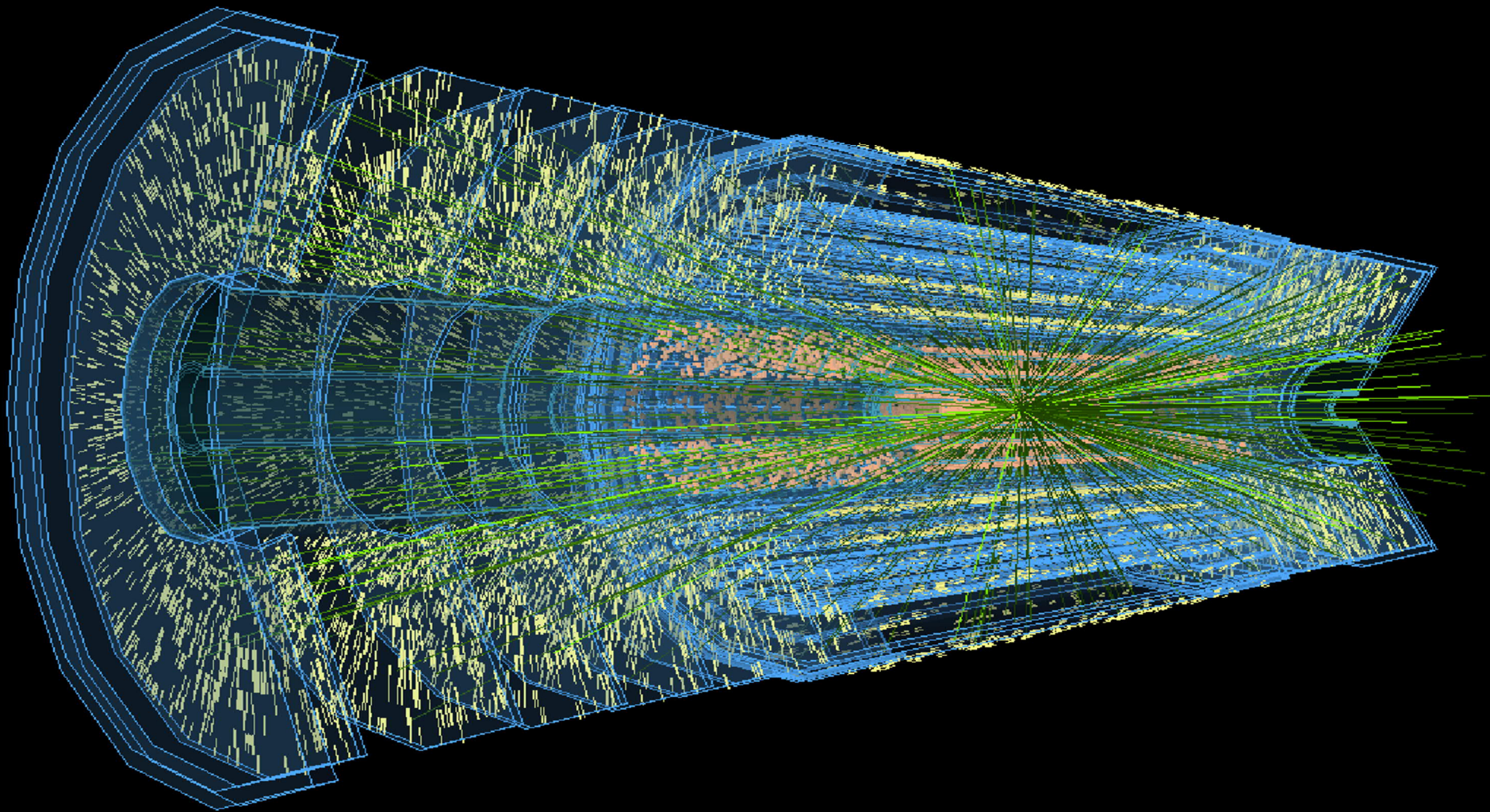
## HL-LHC TECHNICAL EQUIPMENT:



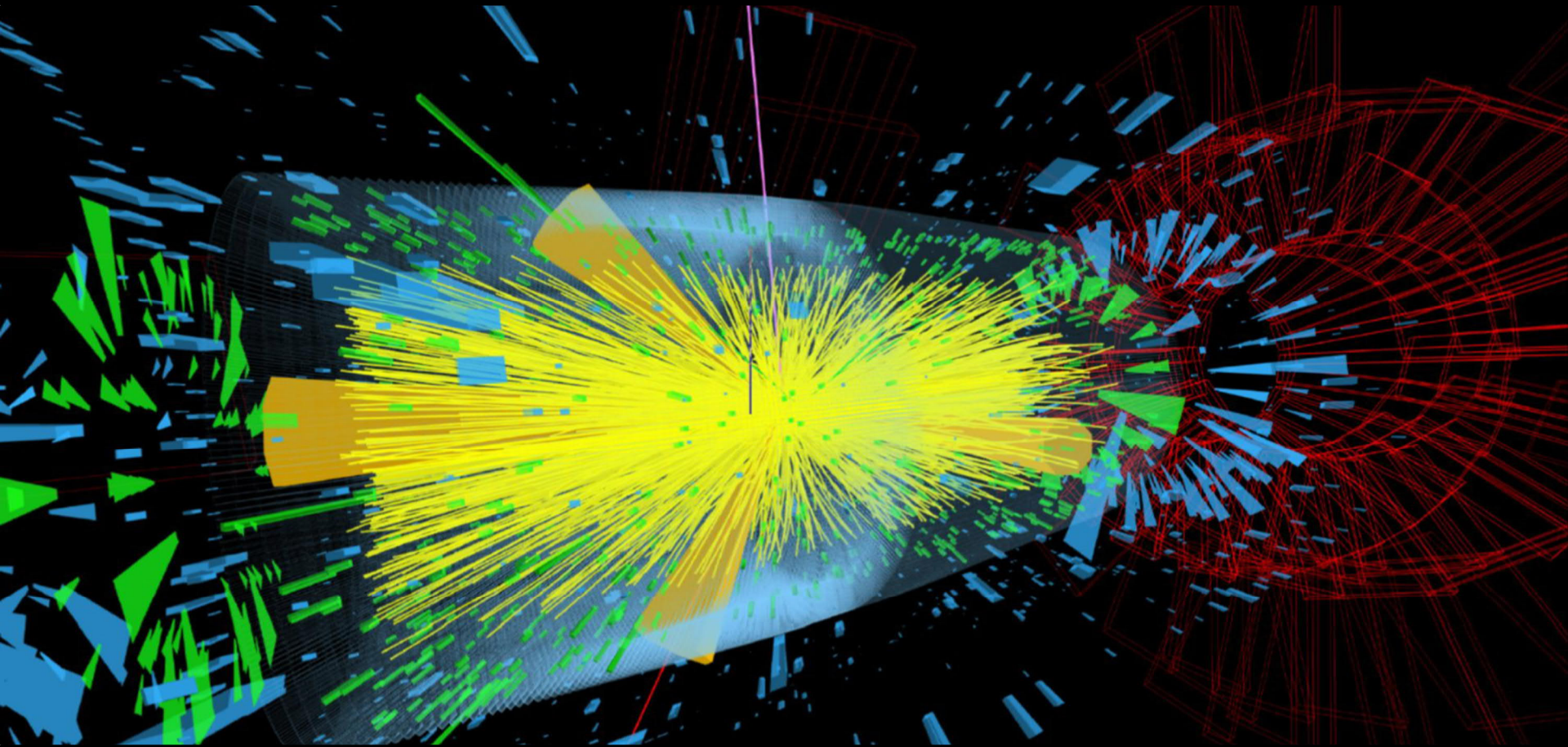
## HL-LHC CIVIL ENGINEERING:













# HL-LHC and friends

- High Luminosity LHC is not alone in the current arena of large scientific collaborations – especially if we look into Astronomy

- Square Kilometer Array (**SKA**)
- Cherenkov Telescope Array (**CTA**)
- NASA James Webb Space Telescope
- ESA Euclid “3D” Telescope
- Etc...

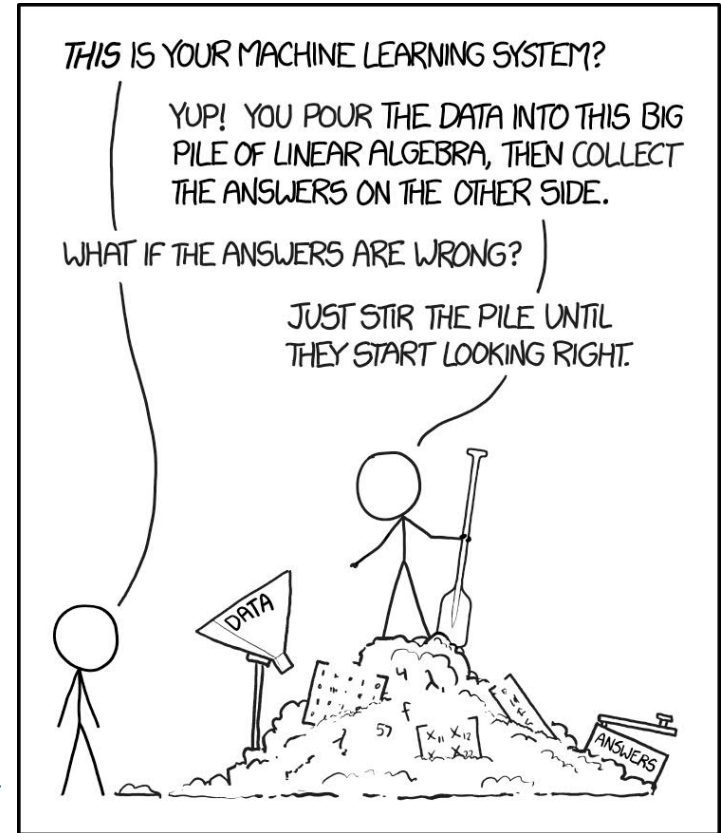
**Large discovery potential, beyond what particle colliders can do!**

- Opportunity for new **synergies** around ML techniques
  - **LIGO/Virgo**: automatic GW signal detection and alerting in **extremely low SNR regime**



# Machine Learning Traps...

This was quoted at the  
CERN Academic Training on  
Machine Learning...



<https://xkcd.com/1838>, May 2017



# GPT Traps...

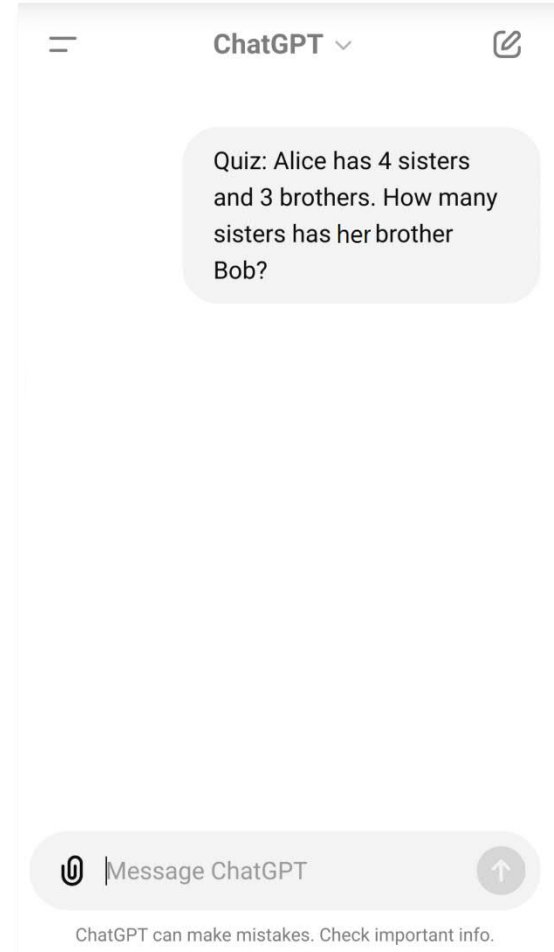
Similarly, Transformers may make mistakes when brought outside of their “comfort zone”.

And worse, they are designed to **always** answer: they very rarely say “I don’t know”

Some LLMs provide a confidence value of their answer

Trivial ad-hoc questions are utterly failed...

<https://arxiv.org/abs/2406.02061>, July 2024



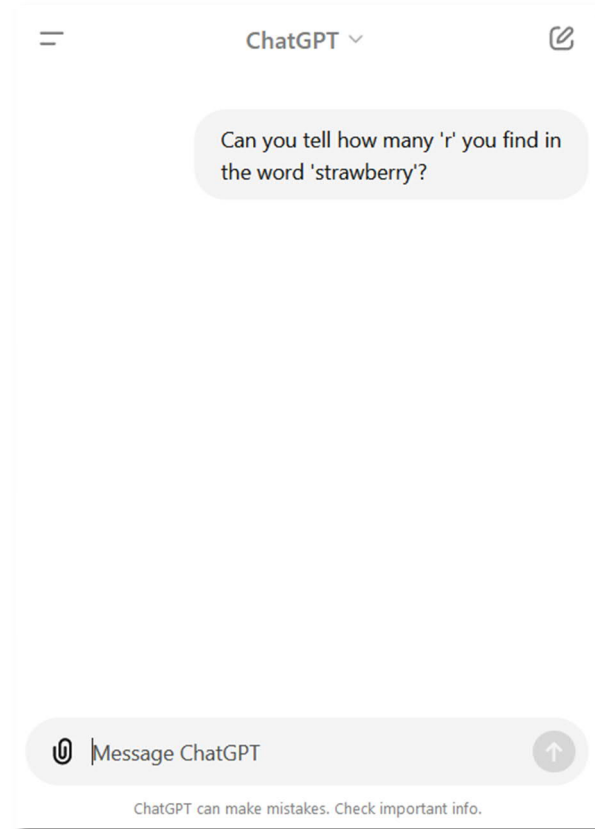
# GPT Traps...

An emerging research area attempts at measuring the “reasoning” abilities of Transformers, beyond their intrinsic “encyclopedic” capabilities.

Specific benchmarks show the remarkable gap when comparing to human intelligence...



Rank	Model	Score	Company
-	Human (avg)	92%	n/a
1st	Claude 3.5 Sonnet	27%	Anthropic
2nd	GPT-4 Turbo-Preview	26%	OpenAI

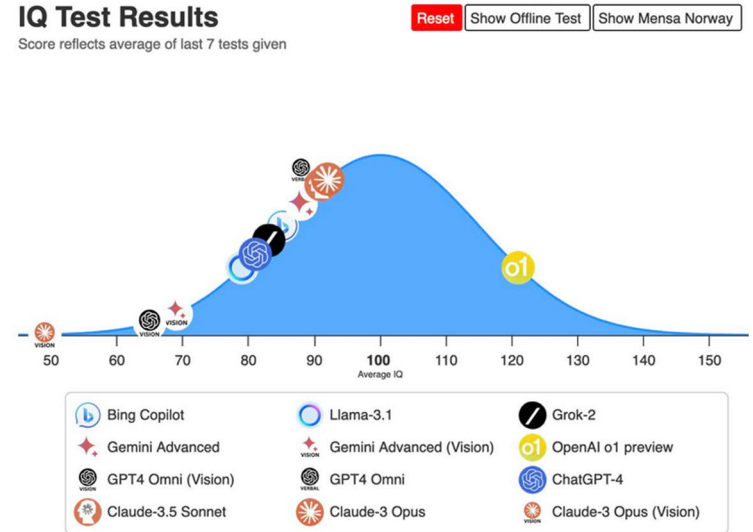


<https://simple-bench.com/index.html>



# The future is bright?

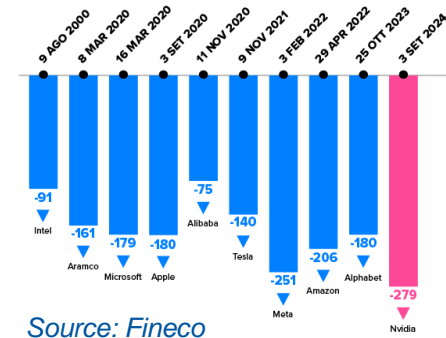
- Of course, the race is open to improve LLMs and achieve **GenAI**, or **General Artificial Intelligence**
- Most recent results seem promising
- But... *Quid est intelligentia?* (cf. Pilate)



<https://trackingai.org/IQ>

# Opportunities and Risks...

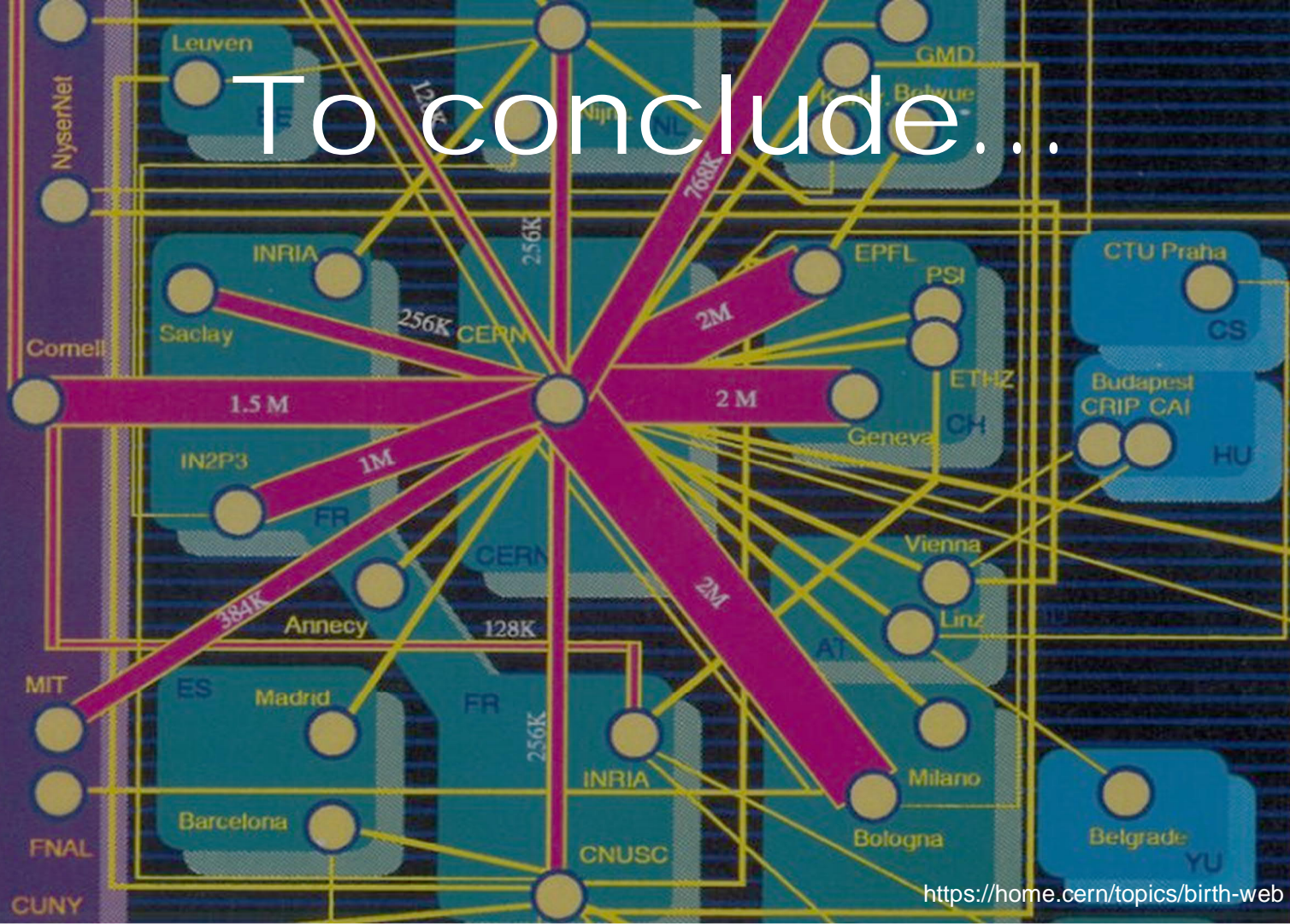
- **Data Science** is a popular career path, crossing the boundaries between **Computer Science**, **Physics** and **Statistics**
  - Fundamental science and engineering remain the pillars to understand technology!
- Big Data and Machine Learning demonstrate **data's ever-growing value**, especially when dealing with **personal data**
  - In **2023**, **7** out of the **top 10** world-largest companies by capitalization – dubbed *The Magnificent Seven* – are entirely **based on the Data economy**
  - At **13.7 T\$**, they compare with the **GDP of Germany + UK + France + Italy!**
    - But with a hugely larger *market volatility*...



Source: Fineco



# To conclude...



# From CERN to the world

- Fundamental Science always pushed technology boundaries, with large returns on investments
- For computing, CERN R&D led for instance to:
  - Invention of the Web (1989)
    - Key contribution to the Internet infrastructure
    - **80% of the total European** Internet traffic going through CERN in the late 1980s
  - Touch screens (1972)
    - Super Proton Synchrotron control system team required complex controls and developed capacitive multi-touch screens
    - It was based on **open standards** and moved into industry



*...mmm... web + touch-screen: what do you have in your pocket/hands?*



# CERN-IT: pushing boundaries

- CERN-IT impact on society through computing:
  - Need for collaboration tools for Global Science led to invent the **World Wide Web**
  - Need for collaboration of computing resources for the Global LHC led to adopt **Grid Computing**, pioneering the concept of **Computing Clouds**
- From Open Source to Open Access to science
  - Need for sharing the results had led CERN to pave the way to open access to documents and now data: **CERN Open Data Portal**
- **Openlab**
  - *Public-private partnership to accelerate the development of cutting-edge solutions for the worldwide LHC community and wider scientific research*
    - Many big IT players involved, including (in alphabetic order) **Google, IBM, Intel, Microsoft, Oracle, ...**
    - Large student internship programme



# Take-away #2

- **Fundamental** Science continues to be main inspiration for **revolutionary** ideas, due to revolutionary needs
  - Industry has well defined offer and demand. We do not. **This is the key for innovation.**
- IT industry has **globally** evolved **beyond our scale**
  - Big Data analysis techniques gaining more and more momentum
    - **But there's no silver bullet ! We must understand the answer !**
  - The role of **Open Source** in software development is more and more crucial as scientific collaborations get larger





**Thanks for your attention! Questions?**



Accélérateur de science

[Giuseppe.LoPresti@cern.ch](mailto:Giuseppe.LoPresti@cern.ch)

[www.linkedin.com/in/giuseppelopresti](https://www.linkedin.com/in/giuseppelopresti)