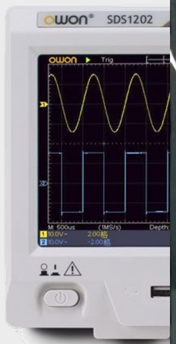
# Transformació Digital: introducció pràctica des de l'Anàlisi de Dades fins a la Intel·ligència Artificial

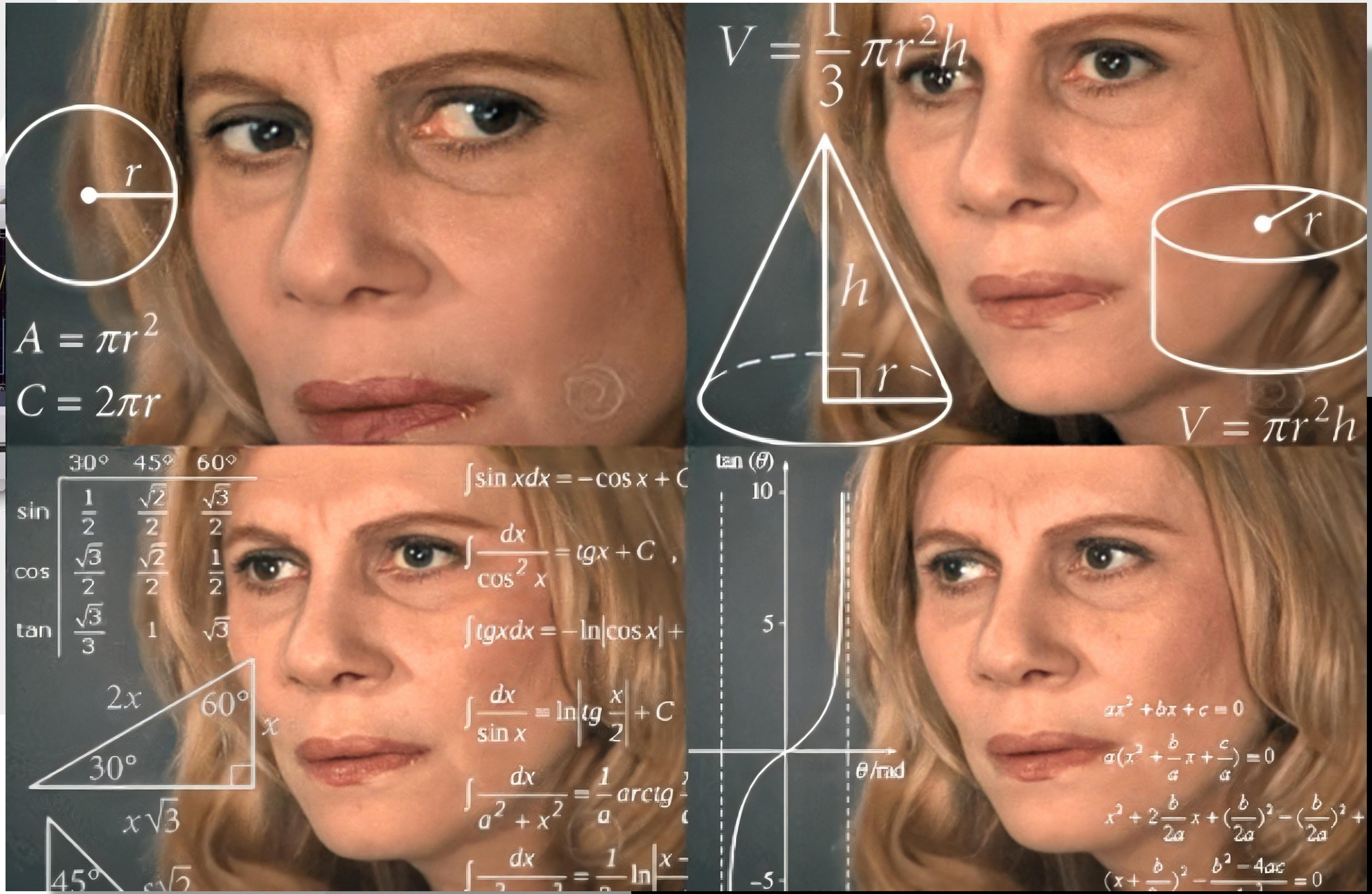
jornada TIC 2023, "Intel·ligentment TIC"

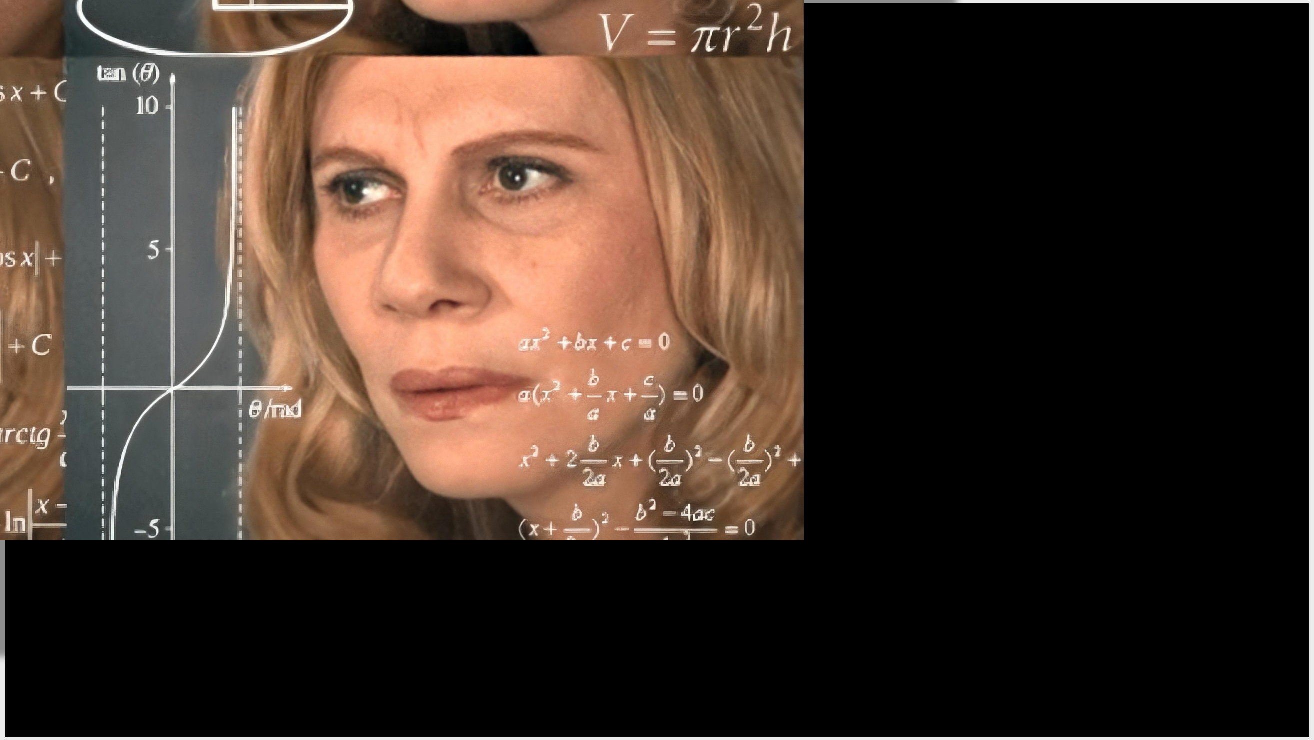


@fcoalberto

https://github.com/elyal2/UPC2023

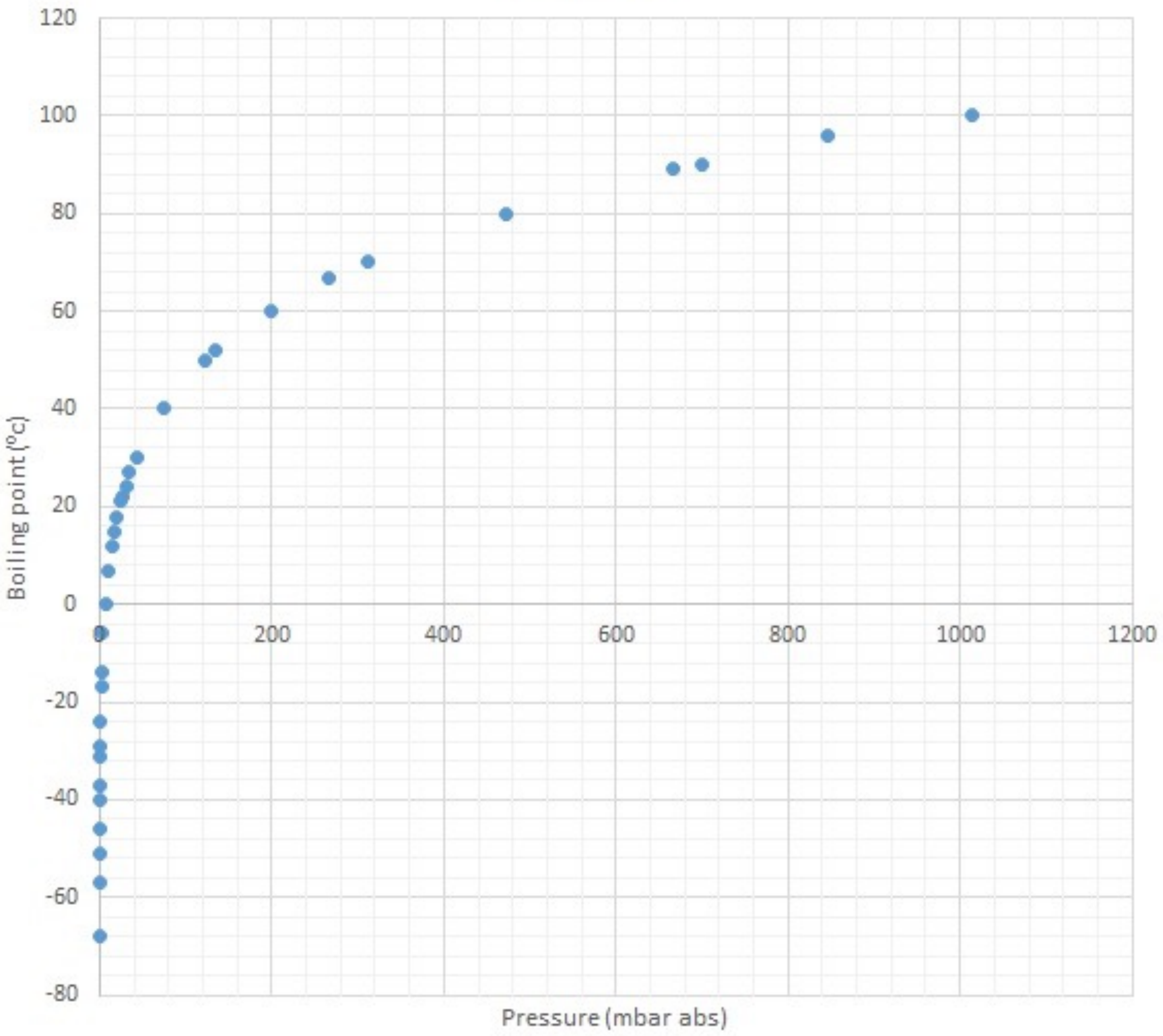


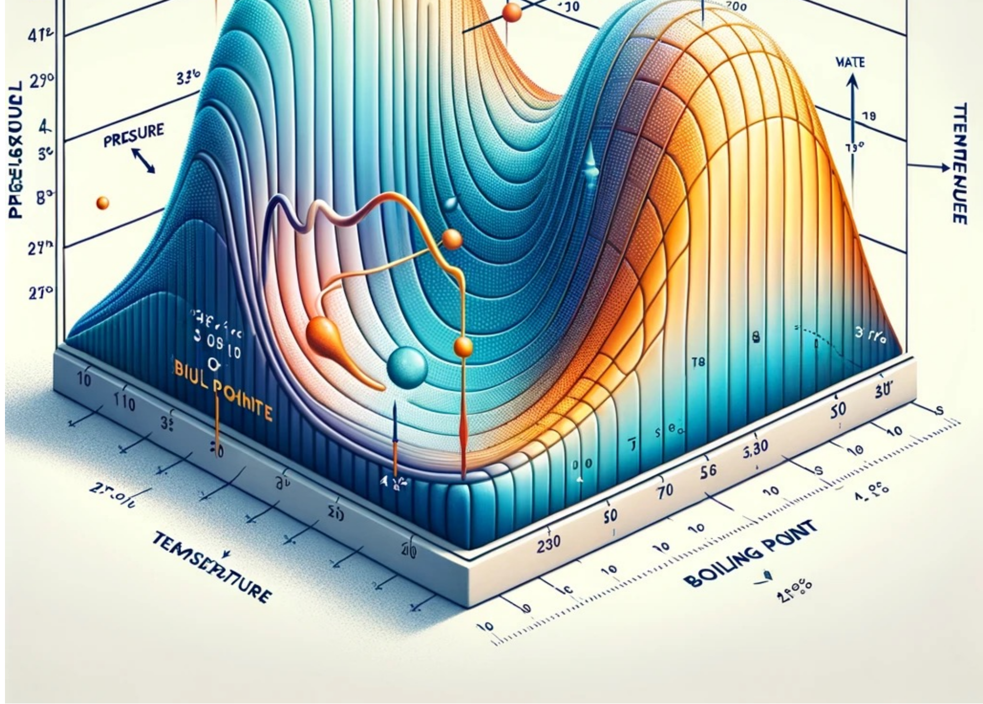



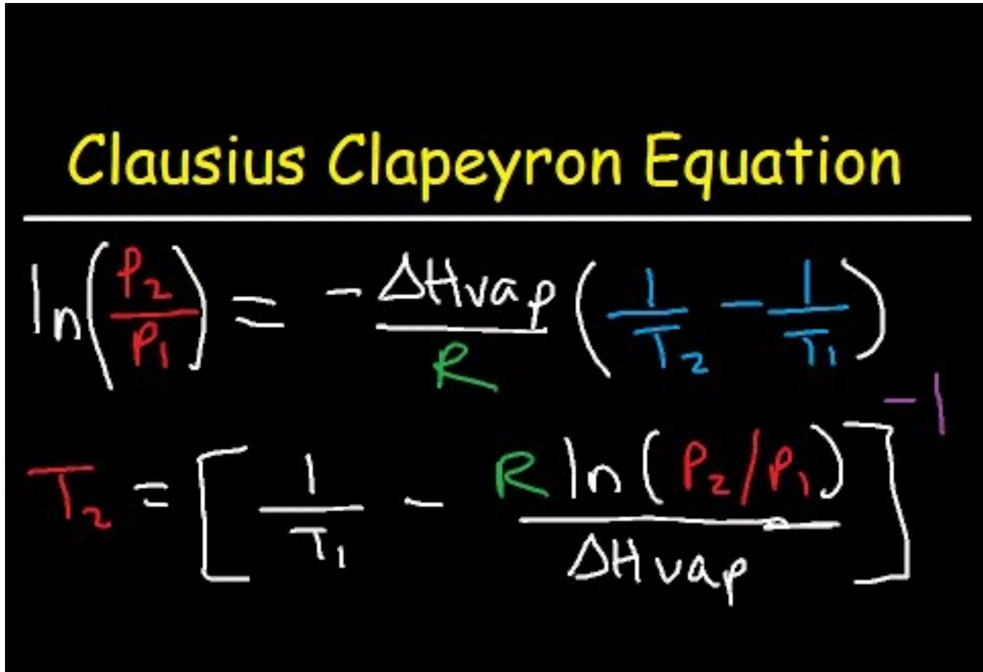



https://oscilloscopemusic.com/software/

# Water boiling point (ºc) as a function of absolute pressure (mbar abs)



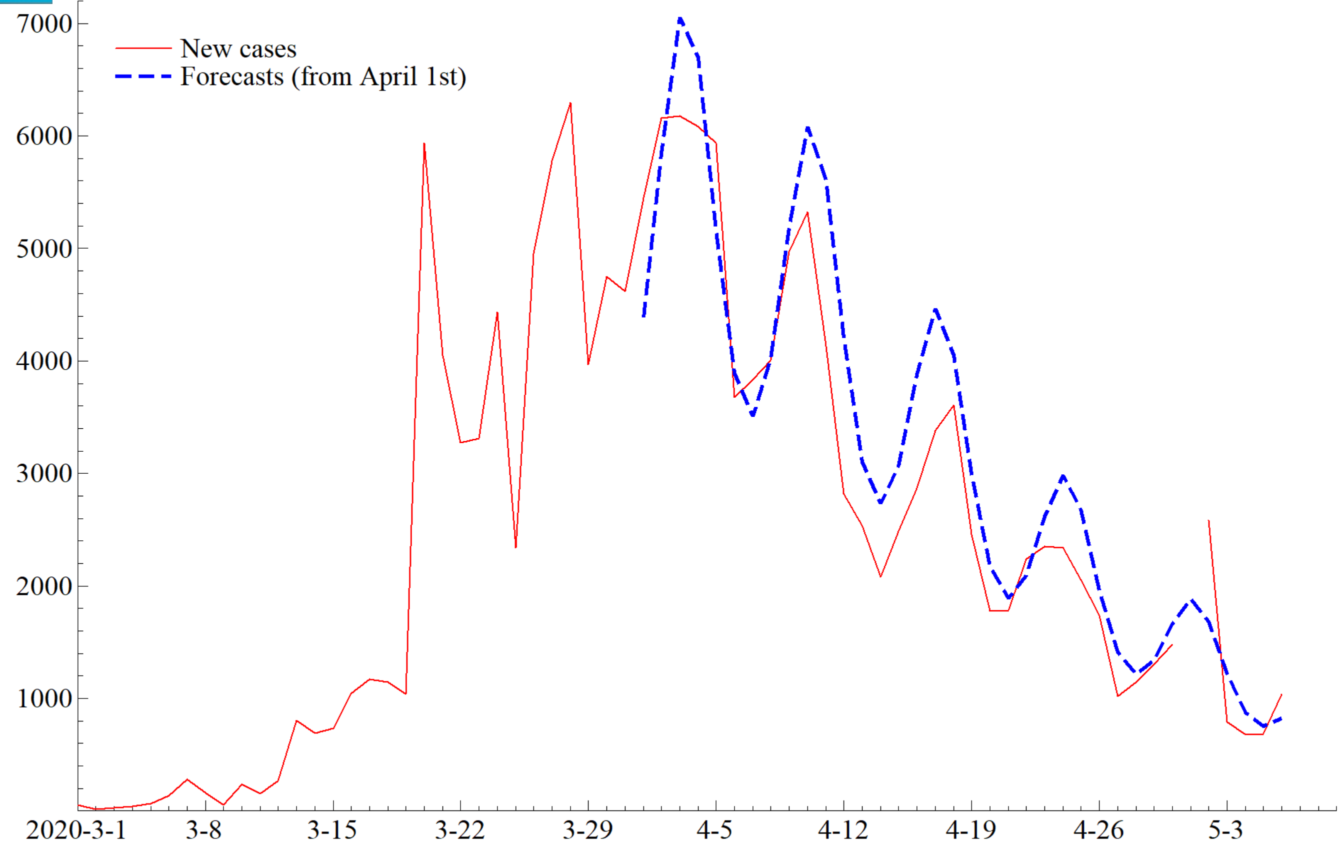






$$\ln\left(\frac{P_2}{P_1}\right) = -\frac{\Delta H vap}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right)$$

$$T_2 = \left[\frac{1}{T_1} - \frac{R\ln(P_2/P_1)}{\Delta H vap}\right]^{-1}$$

https://www.youtube.com/watch?v=TkwXa7Cvfr8




Forecasts (dashed - blue) for new cases in Germany using data (solid - red) up to March 31st. There was no observation reported for May 1st. The observation for May 2nd can be regarded as the sum of observations for May 1st and 2nd.

https://hdsr.mitpress.mit.edu/pub/ozgjx0yn/release/4

## 2. Growth Curves

Let $\mu(t) \geq 0$ be a monotonically increasing function defined over the real line. The rate of change or 'incidence curve' is $d\mu(t)/dt \geq 0$. The generalized logistic is

(2.1) $$\mu(t) = \overline{\mu}/(1+(\gamma_0/\kappa)e^{-\gamma t})^{\kappa}, \quad \gamma_0, \gamma, \kappa > 0,$$

where $\gamma$ is a growth rate parameter. The parameter $\kappa$ must be positive for there to be an upper asymptote; allowing $\kappa$ to be negative gives the class of general modified exponential (GME) growth curves. The logistic is obtained by setting $\kappa = 1$, while letting $\kappa \to \infty$ yields the Gompertz curve. When $\gamma_0$ is determined by the value of the curve at $t = 0$, it is

(2.2) $$\gamma_0 = \kappa\left[(\overline{\mu}/\mu(0))^{1/\kappa} - 1\right].$$

Differentiation yields

(2.3) $$\ln d\mu(t)/dt = \rho \ln \mu(t) + \delta - \gamma t,$$

where $\delta = \ln(\gamma_0\gamma/\overline{\mu}^{\,1/\kappa})$ and $\rho = (\kappa+1)/\kappa$, so $0 < \kappa < \infty$ implies $1 < \rho < \infty$. Alternatively, because $d\mu(t)/dt = g(t)\mu(t)$,

(2.4) $$\ln g(t) = (\rho - 1)\ln \mu(t) + \delta - \gamma t,$$

where $g(t)$ is the growth rate of $\mu(t)$. Note that $\rho - 1 = 1/\kappa$.

The generalized logistic differential equation implied by Equation 2.1 is

(2.5) $$\frac{d\mu(t)}{dt} = \gamma\kappa\left[1 - \left(\frac{\mu(t)}{\overline{\mu}}\right)^{1/\kappa}\right]\mu(t).$$

The term in square brackets is less than one and tends to zero as $\mu(t) \to \overline{\mu}$. The growth rate implied by Equation 2.5 is

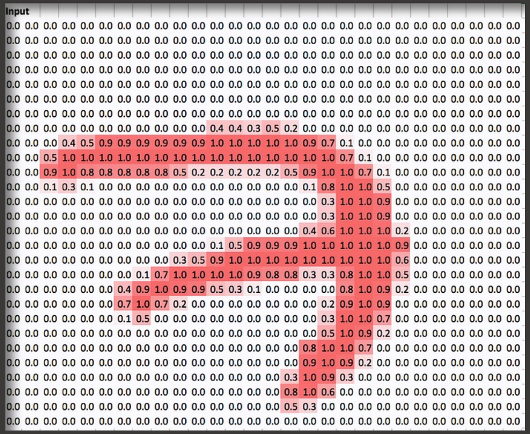
(2.6) $$g(t) = \gamma\kappa\left[1 - \left(\frac{\mu(t)}{\overline{\mu}}\right)^{1/\kappa}\right].$$

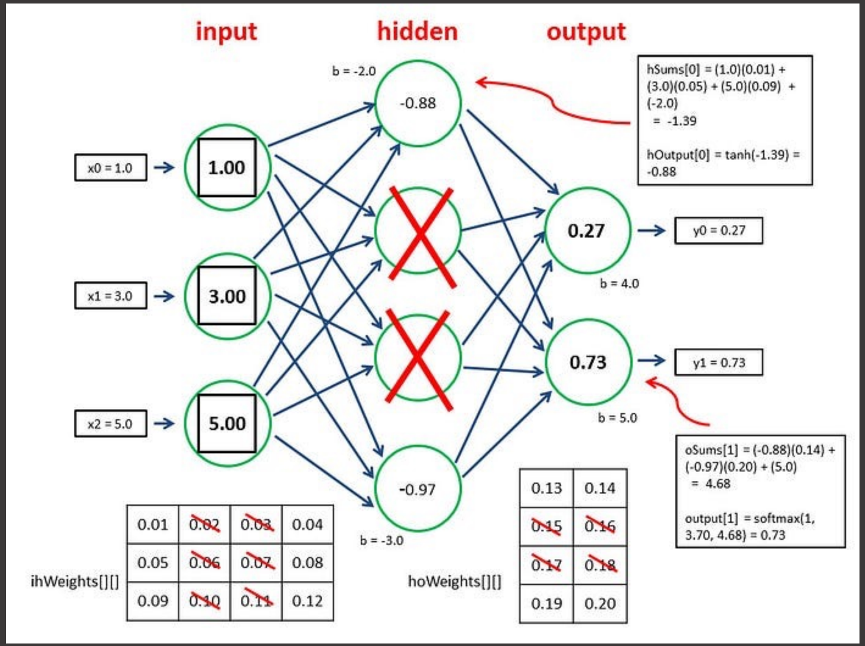
When $\kappa = 1$, Equation 2.5 is a Riccati equation.


There is Certainty, Doubt, and Probability

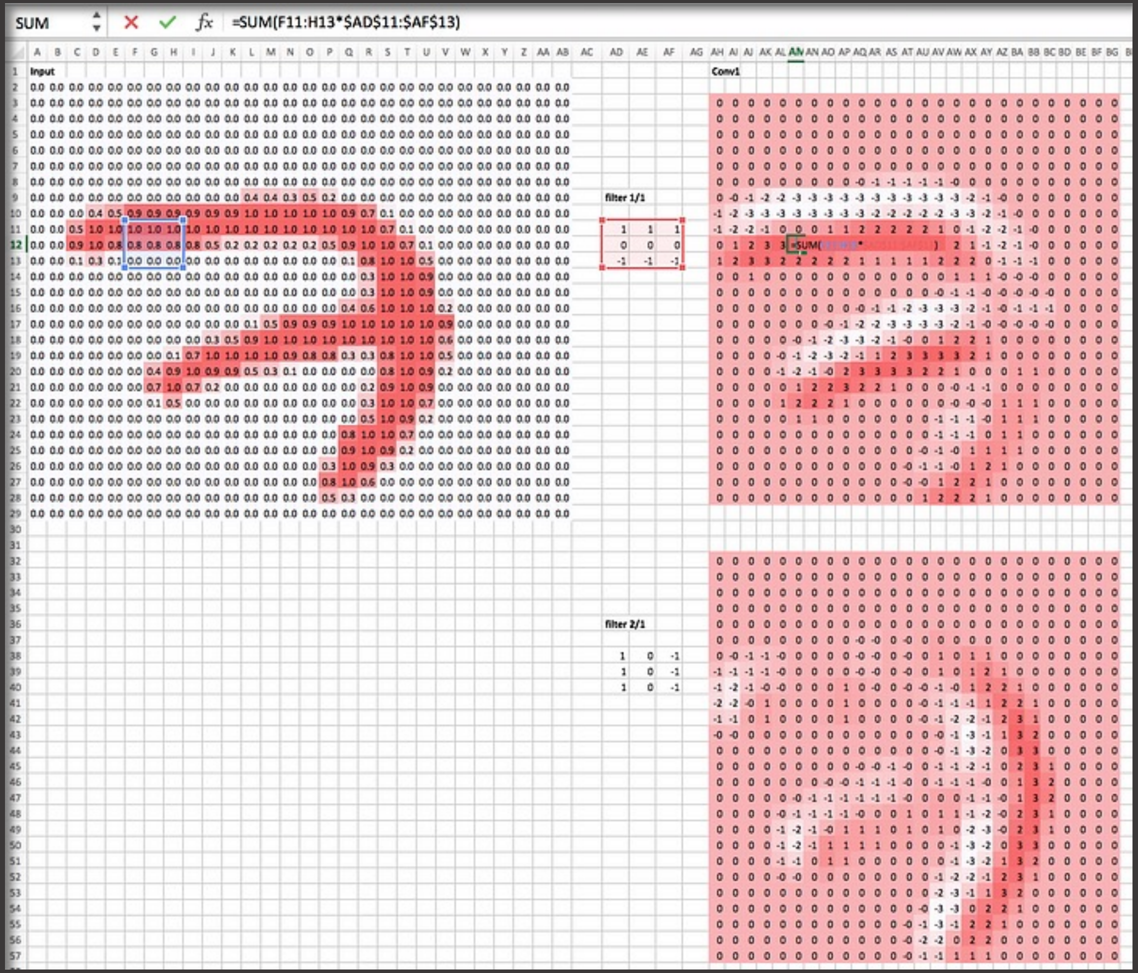
... and then there is 95% probability

# = 0
# = .1
# = .3
# = .5
# = .7
# = .9
# = 1

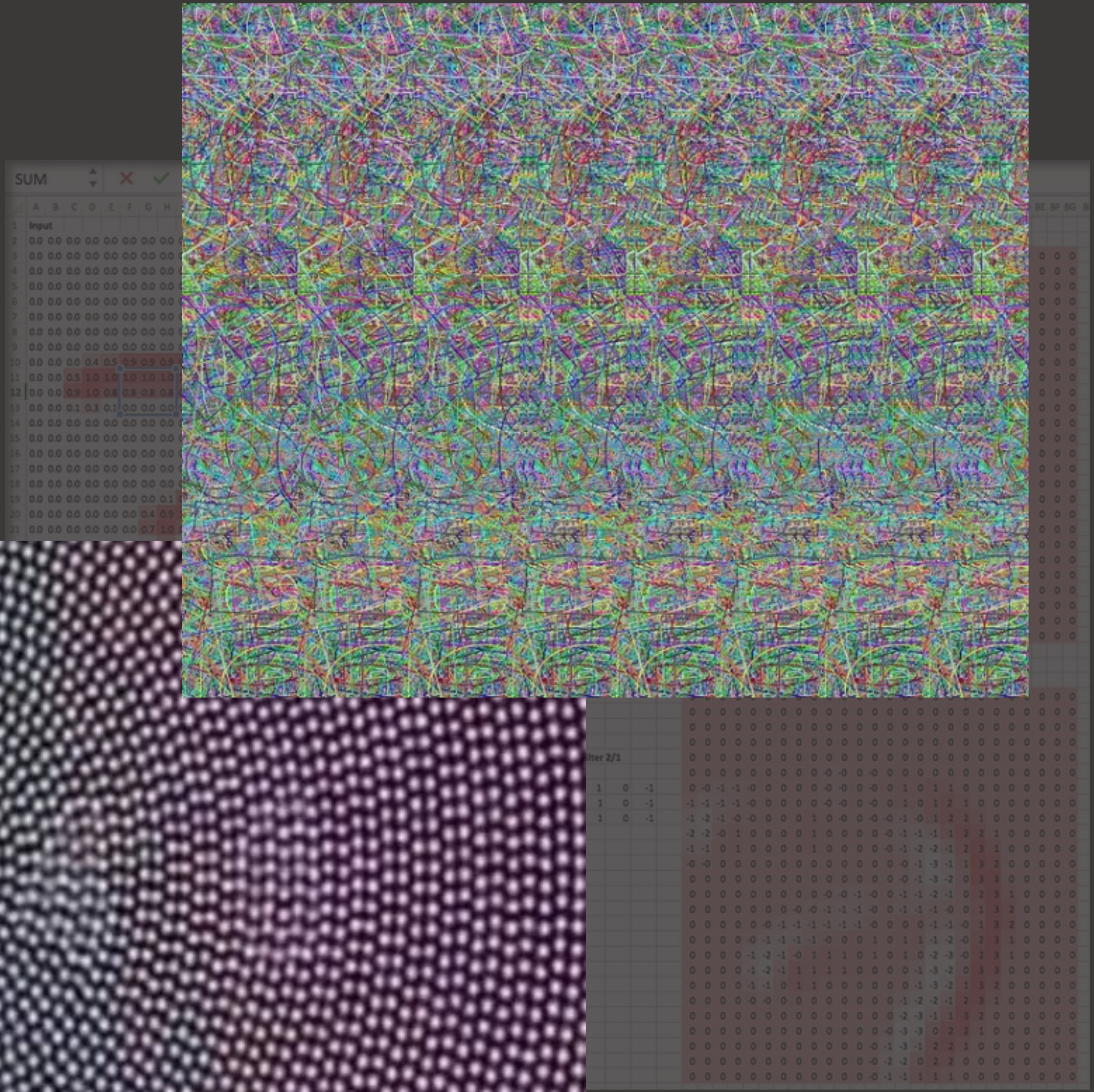


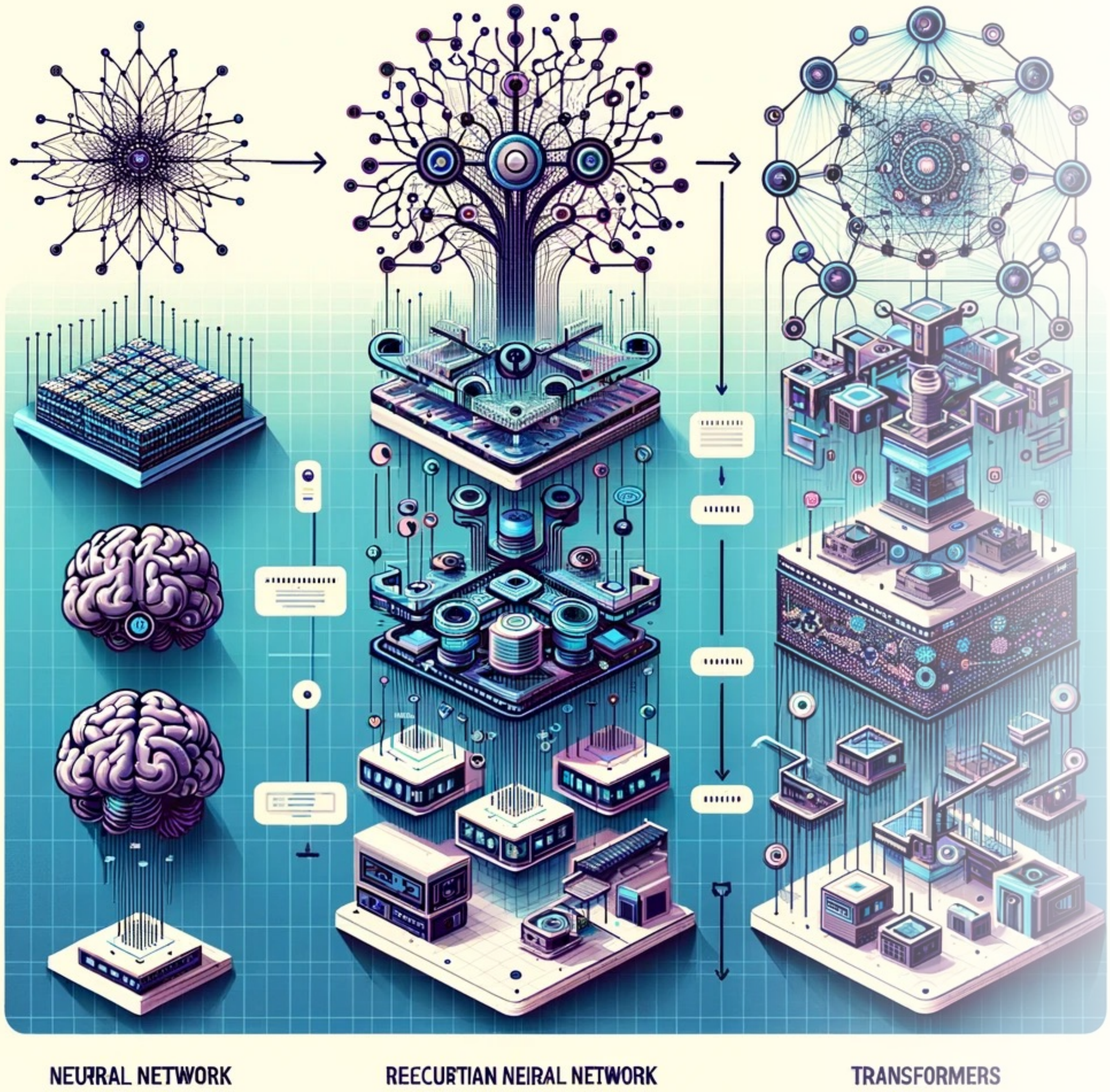

https://medium.com/@prathammodi001/convolutional-neural-networks-for-dummies-a-step-by-step-cnn-tutorial-e68f464d608f




https://towardsdatascience.com/understanding-convolutions-using-excel-886ca0a964b7


olutions-using-excel-886ca0a964b7




**1950 - Neural Networks (NNs)**: The foundation of modern AI, these basic structures consist of interconnected nodes (neurons) that process data in layers, enabling pattern recognition and decision-making.

**1980 (Fukushima) - Recurrent Neural Networks (RNNs)**: An advancement over NNs, RNNs are designed to handle sequential data. They incorporate loops within their architecture, allowing information to persist, which is vital for tasks like language modelling and time series analysis.

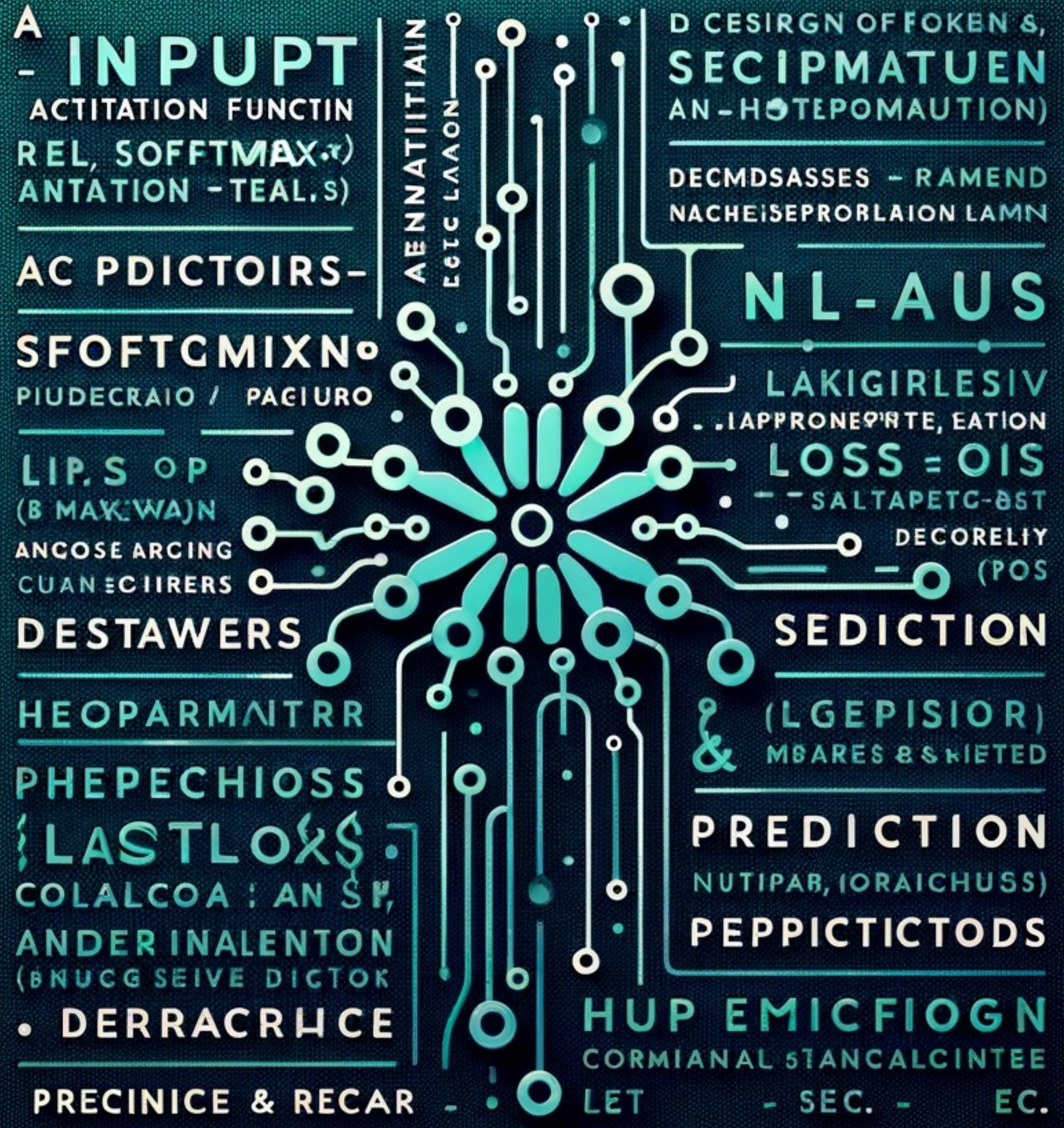
**1989 (Yann LeCun) - Convolutional Neural Networks (CNNs)**: CNNs are particularly structured for processing data that has a grid-like topology. This makes them highly efficient for tasks involving images (which can be viewed as 2D grids of pixels). They use layers to filter inputs for useful information, reducing the dimensions of the data while preserving essential features. This is followed by more layers that further downsample the data.

**2017 (Google) - Transformers**: The latest breakthrough, transformers, move beyond sequential data processing constraints of RNNs. They employ self-attention mechanisms, efficiently handling large sets of data and excelling in complex tasks like natural language processing, significantly improving speed and accuracy.

https://projector.tensorflow.org/

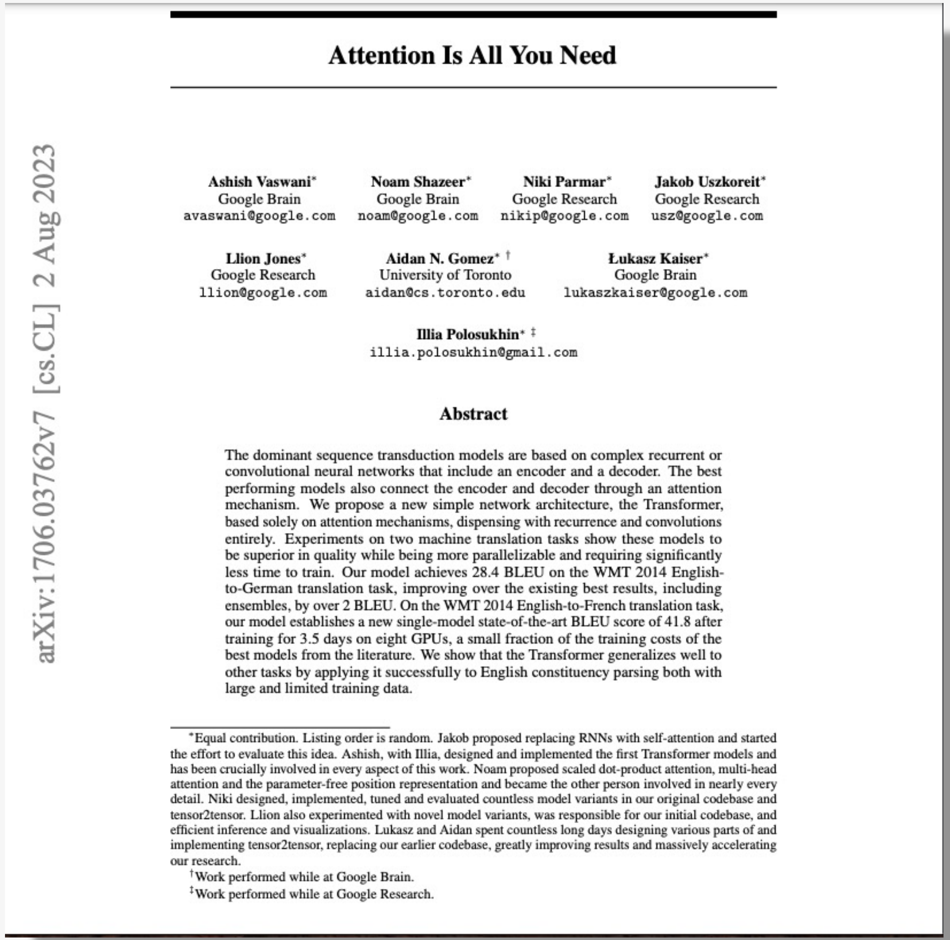
https://playground.tensorflow.org/

INPUT
ACTIVATION FUNCTION
PARAMETERS
HYPERPARAMETERS
STRATEGIES
PREDICTION
LOSS
PRECISION & RECALL
OVERFITTING




https://www.youtube.com/watch?v=TkwXa7Cvfr8

# 2017



## 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

arXiv:1706.03762v7 [cs.CL] 2 Aug 2023

### Abstract

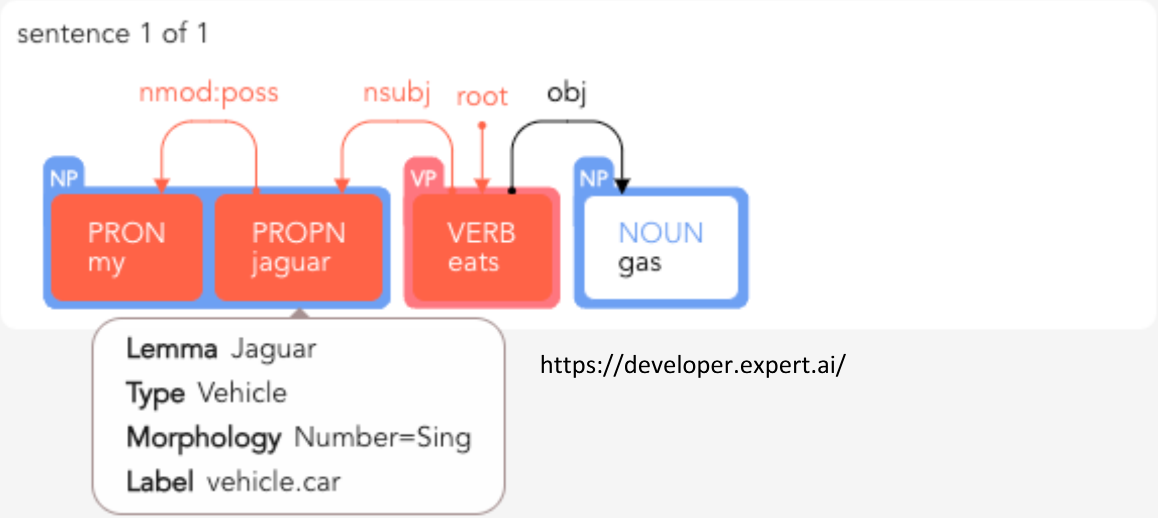
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, 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.

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

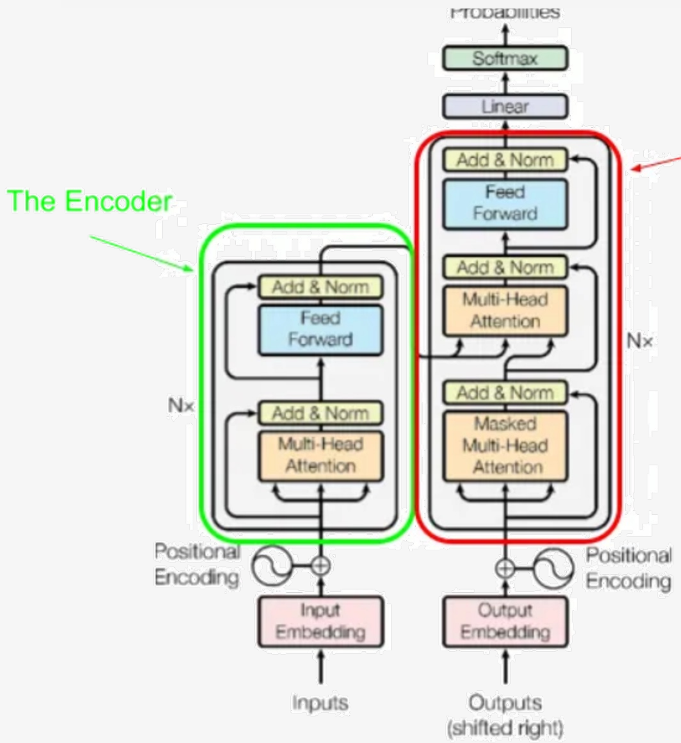
‡Work performed while at Google Research.

https://arxiv.org/pdf/1706.03762.pdf

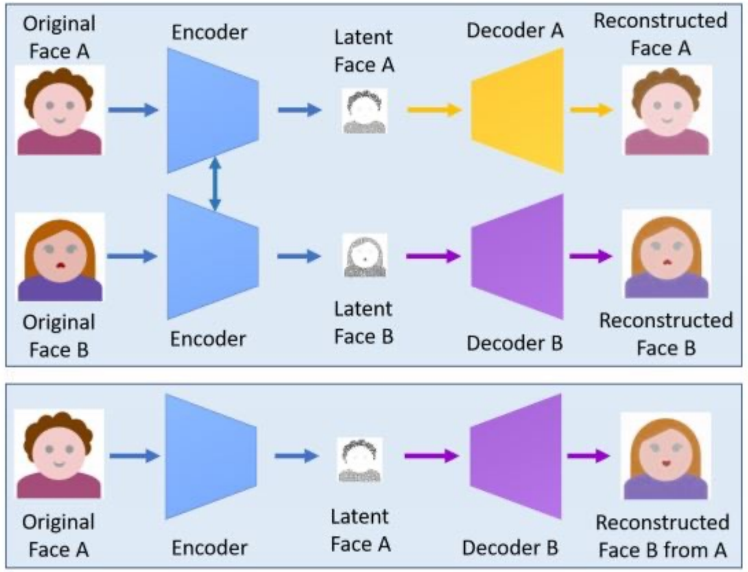



https://developer.expert.ai/

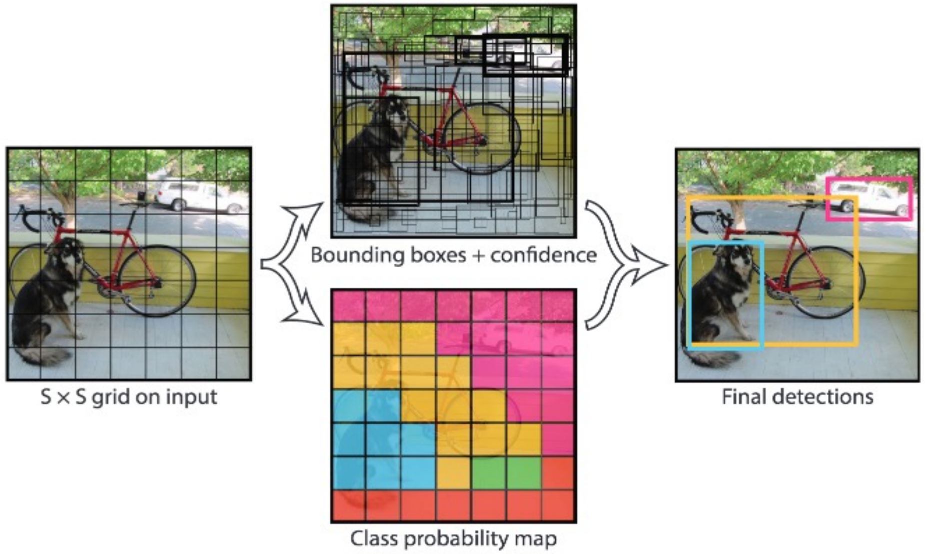
The Decoder



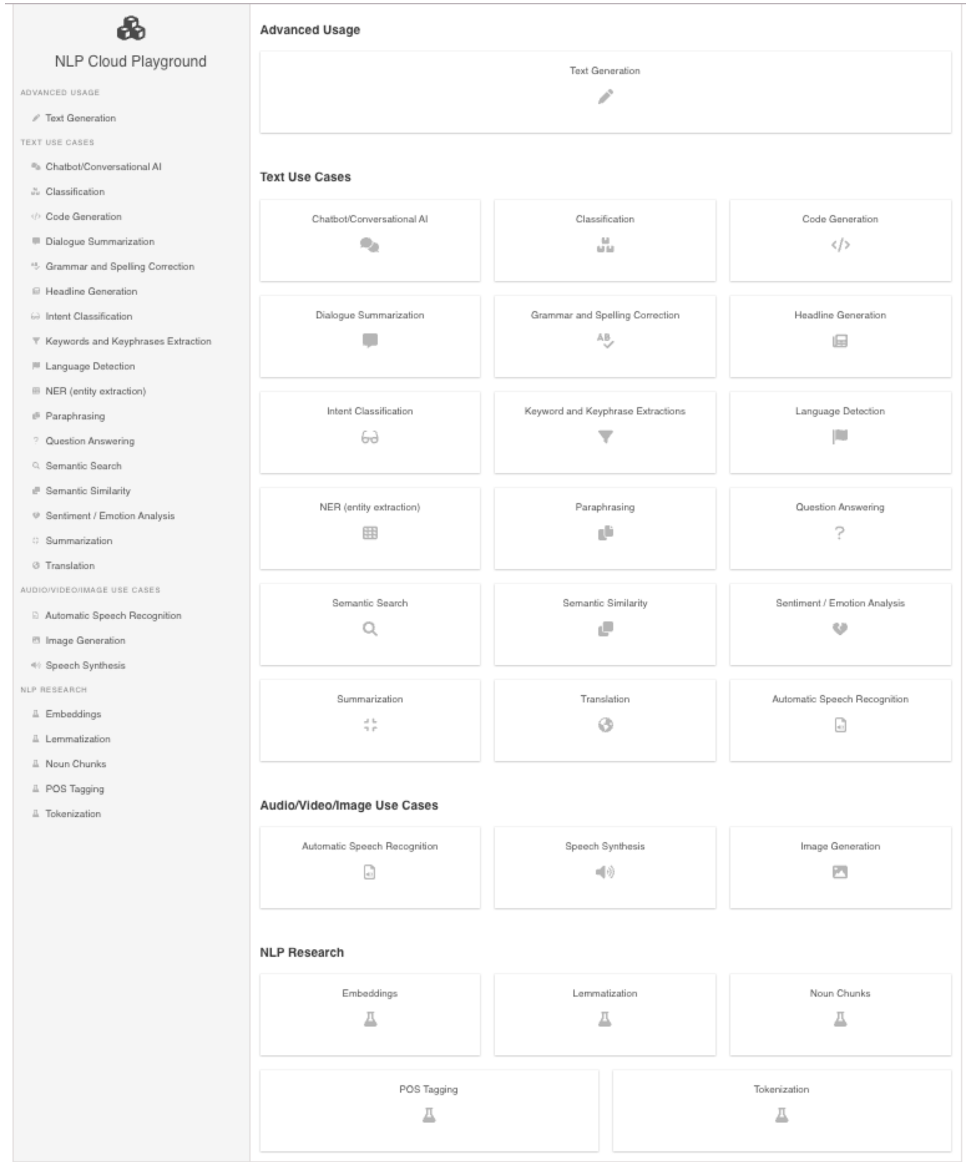

1. **Masked self-attention**: helps to prevent the decoder from generating nonsensical or repetitive text.
2. **Decoder self-attention**: build on what it has already produced and maintain coherence.
3. **Decoder-encoder attention**: context for generating the output sequence.




https://arxiv.org/pdf/1909.11573.pdf




https://cocodataset.org/

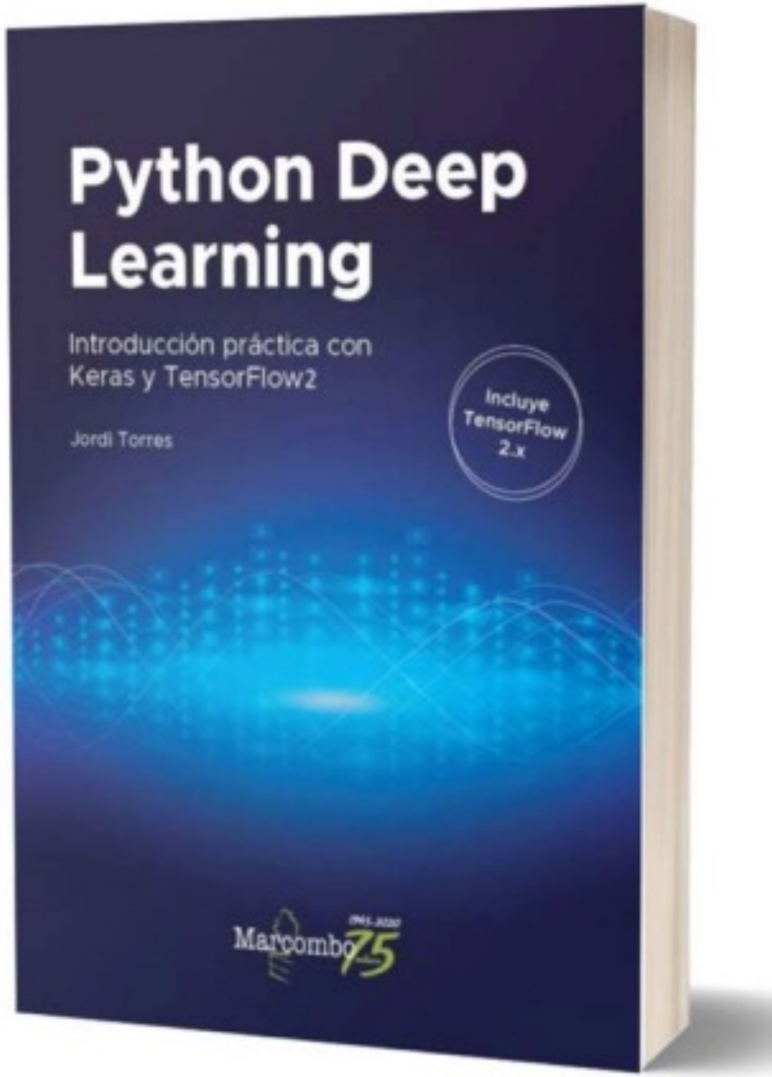



https://nlpcloud.com

Building a large language model (LLM) compared to a traditional model is like quantifying the grains of sand on a beach: where traditional models apply clever formulas for a rough estimate, an LLM embarks on the colossal task of meticulously counting each grain.

Jordi TORRES.AI

https://www.marcombo.com/python-deep-learning-9788426728289/




Barcelona Supercomputing Center
Centro Nacional de Supercomputación

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

https://huggingface.co/spaces/AP123/IllusionDiffusion