Modelos de Deep Learning

Introducción

Universidad ORT Uruguay

18 de Agosto, 2025

"How can computers learn to solve problems without being explicitly programmed?"

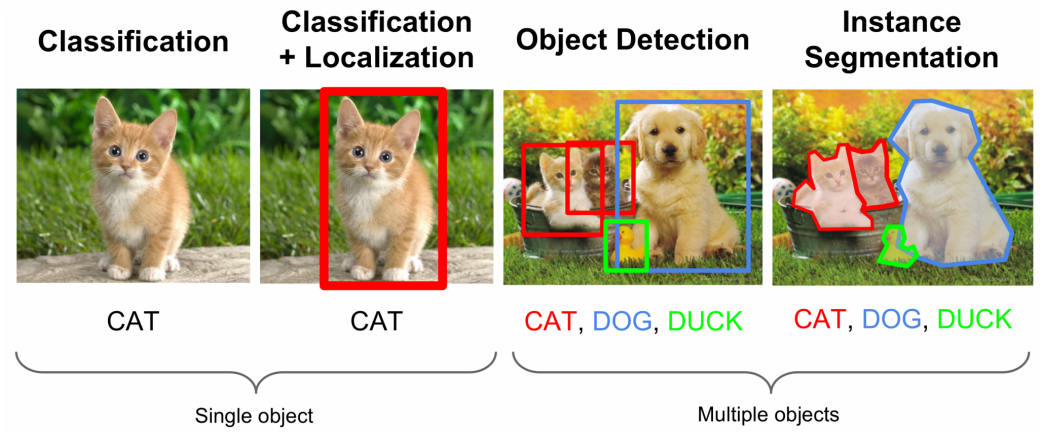
Arthur Samuel (1959)

Definición de Wikipedia:

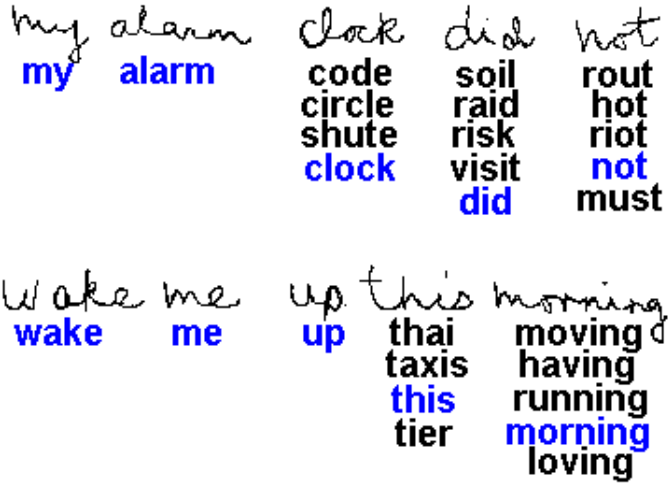
Machine learning is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data and thus perform tasks without explicit instructions.

Lectura recomendada: R. Sutton "The Bitter Lesson" (2019).

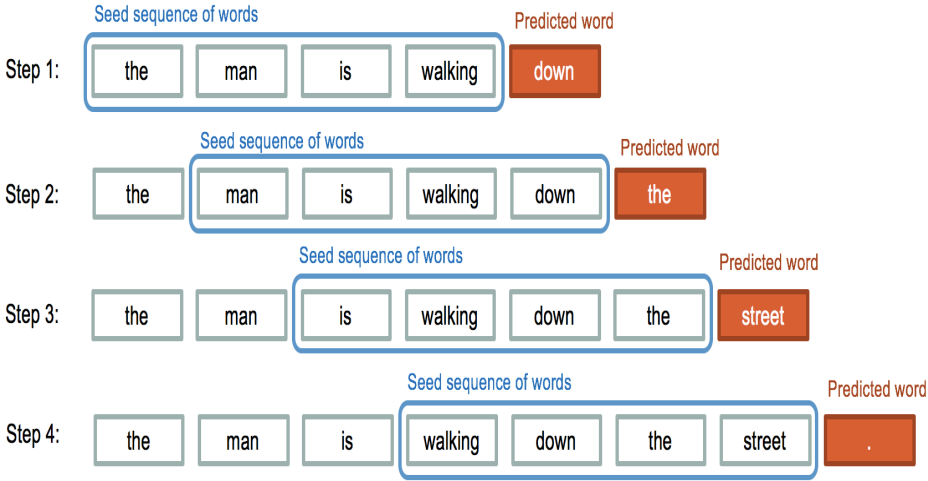
Ejemplo: Computer Vision



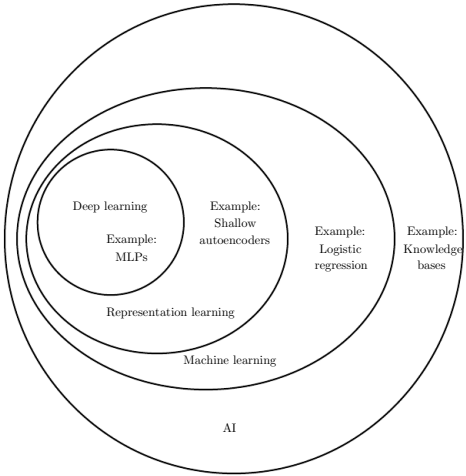
Ejemplo: Handwritten recognition



Ejemplo: Generación de texto



Inteligencia Artificial (AI) - Machine learning (ML)



I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. MIT Press, 2016.

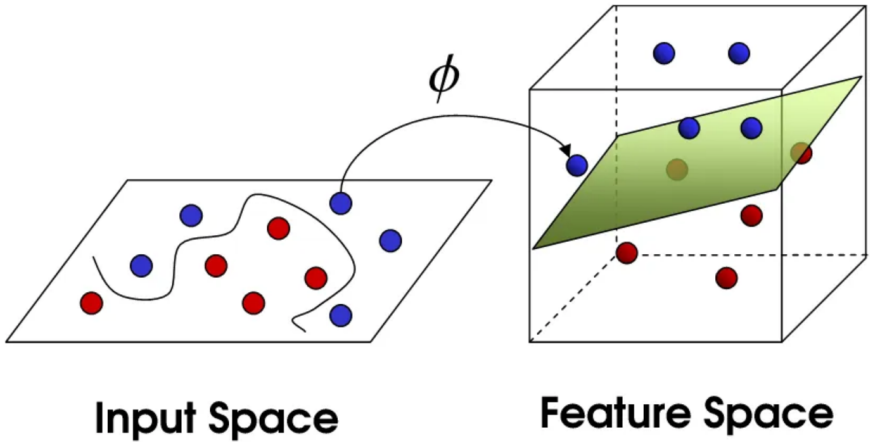
Representación

- Datos
- Hipótesis

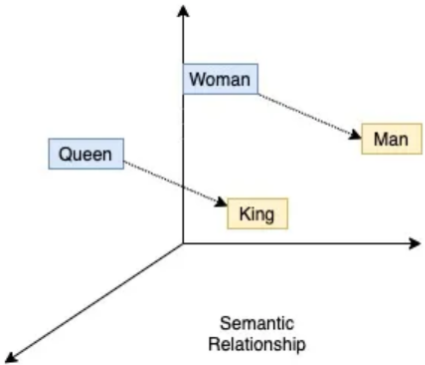
Memorización

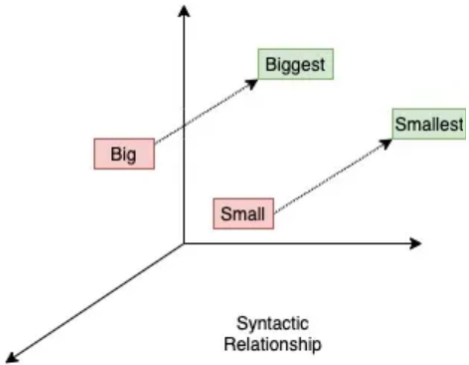
Generalización

Representación: la idea

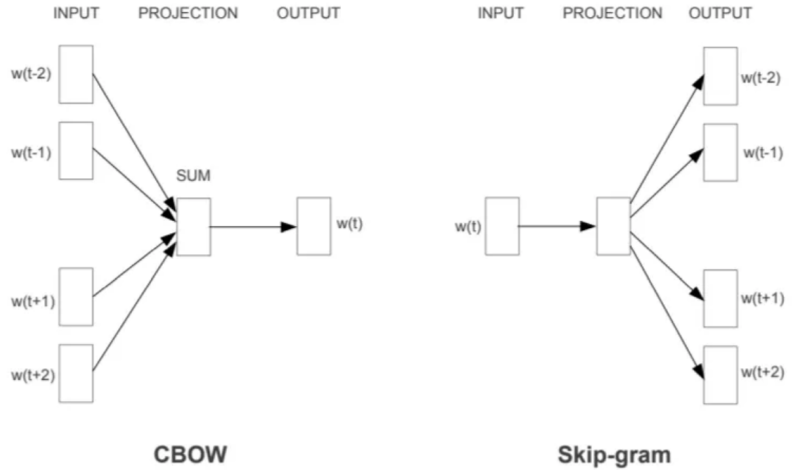


Representación: Word Embeddings



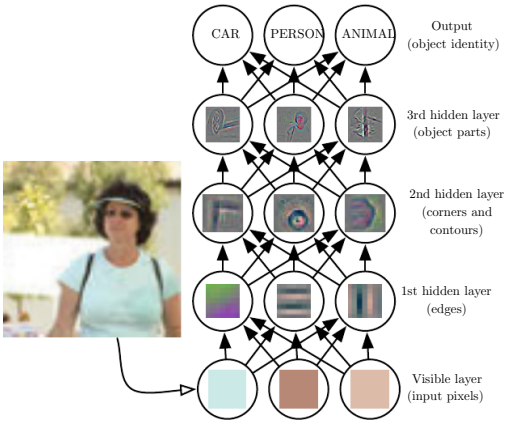


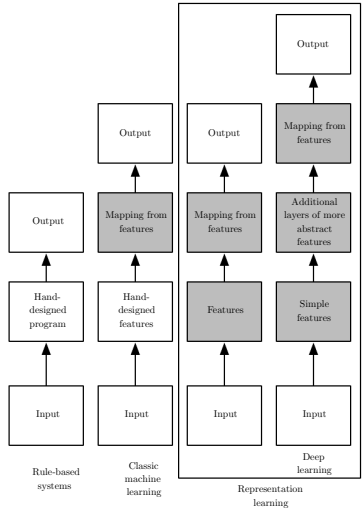
Ejemplo: Word2Vec



Word2Vec: Efficient Estimation of Word Representations in Vector Space (2013)

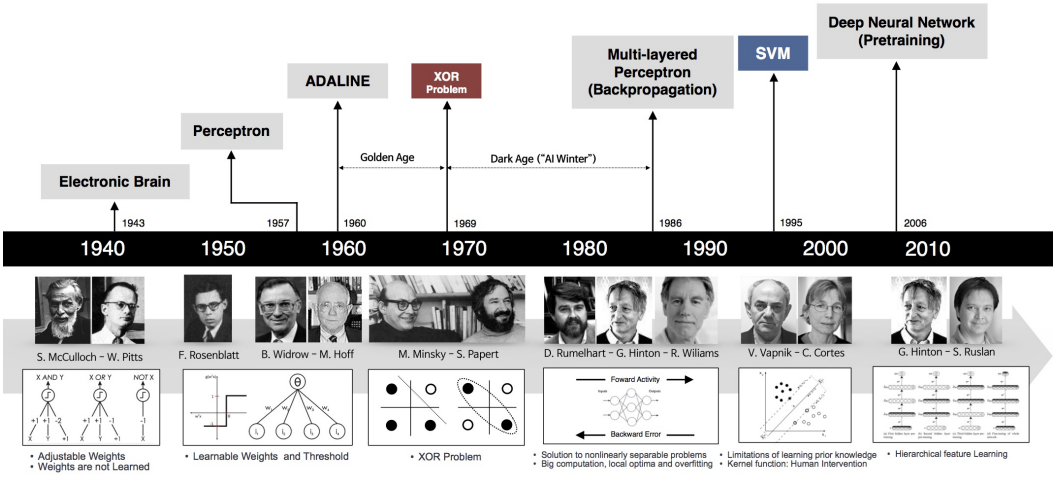
Representación composicional





I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. MIT Press, 2016.

Deep Learning: linea de tiempo

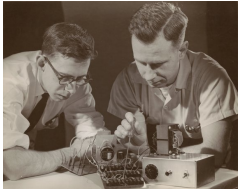


Artificial Neural Networks: Origen

1943 W. McCulloch, W. Pitts.

A Logical Calculus of Ideas Immanent in Nervous Activity

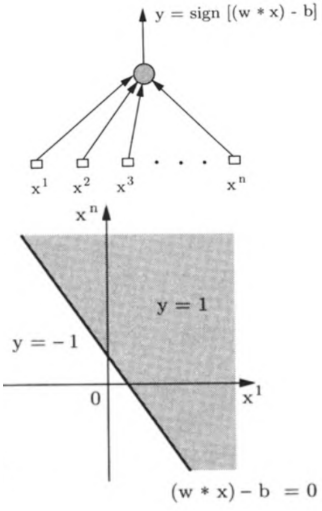
Bull. Math. Biophysics 5 (4): 115-133.



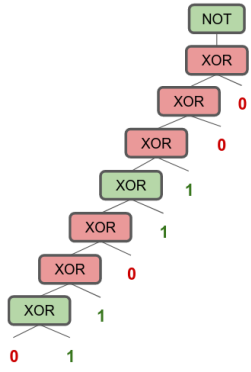
1958 F. Rosenblatt.

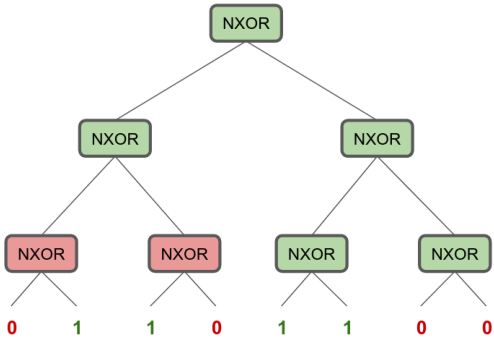
The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain

Psychological Review 65 (6): 386-408.



The XOR Problem and The Parity Problem





The New XOR Problem

Saturated Transformers are Constant-Depth Threshold Circuits

1986 Rumelhart, Hinton & Williams introducen **Backpropagation Algorithm**



Learning representations by back-propagating errors

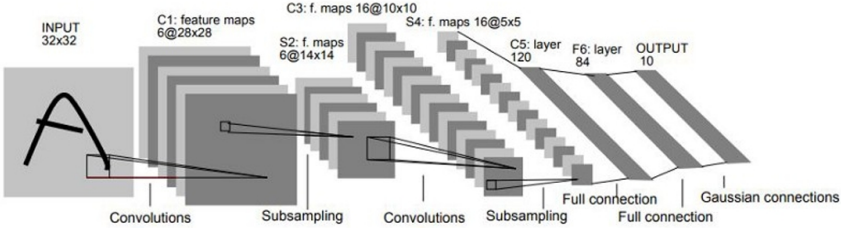
David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].

Nature. 323 (6088): 533-536.

1998 Y. LeCun **Redes Convolucionales**



En Bell Labs, contaba con una red neuronal para el reconocimiento óptico de caracteres (5–7 capas), denominada *LeNet-5*, idéntica en su diseño y entrenamiento a una red neuronal moderna. Sin embargo, datos y capacidad de cómputo limitados, llevaron a un nuevo invierno.

LeCun, Y., *et al.*, 1998. Gradient-based learning applied to document recognition.

Deep Learning

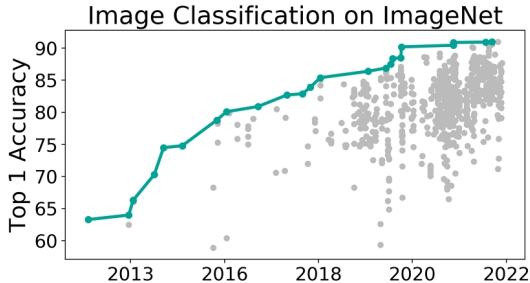
2006 Hinton, G. E., Osindero, S. and Teh, Y.

A fast learning algorithm for deep belief nets.
Neural Computation, 18, pp 1527-1554.

2009 Yoshua Bengio

Learning Deep Architectures for AI.
Foundations and Trends in Machine Learning, 2(1), 1-127.

2012 Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton **AlexNet**



Evolución de la precisión en el *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC). La victoria de 2012 por *AlexNet* (una *LeNet* ligeramente más grande) fue clave para reavivar el interés en las redes neuronales.

ImageNet classification with deep convolutional neural networks.

2014 D. Bahdanau, K. Cho, Y. Bengio **Attention Mechanism on RNNs**

"The models proposed recently for neural machine translation often belong to a family of encoder-decoders and consists of an encoder that encodes a source sentence into a fixed-length vector from which a decoder generates a translation. In this paper, we conjecture that the use of a fixed-length vector is a bottleneck in improving the performance of this basic encoder-decoder architecture, and propose to extend this by allowing a model to automatically (soft-)search for parts of a source sentence that are relevant to predicting a target word, without having to form these parts as a hard segment explicitly. With this new approach, we achieve a translation performance comparable to the existing state-of-the-art phrase-based system on the task of English-to-French translation. Furthermore, qualitative analysis reveals that the (soft-)alignments found by the model agree well with our intuition."

Neural Machine Translation by Jointly Learning to Align and Translate.

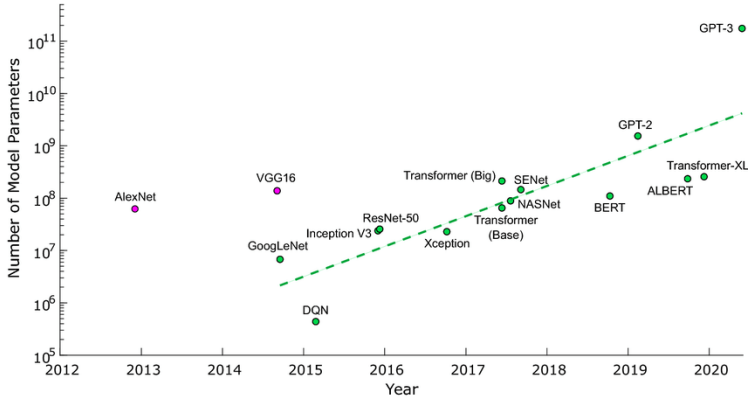
2017 Ashish Vaswani *et al.* **Transformers**

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

La publicación de *Attention Is All You Need* en 2017 introdujo la arquitectura *Transformer*. Su mecanismo de atención auto-regresiva y su capacidad de paralelización impulsaron avances importantes en modelos de gran escala y establecieron las bases de la actual generación de modelos de lenguaje.

Attention Is All You Need.

Deep Learning: Hoy - Cantidad de parámetros

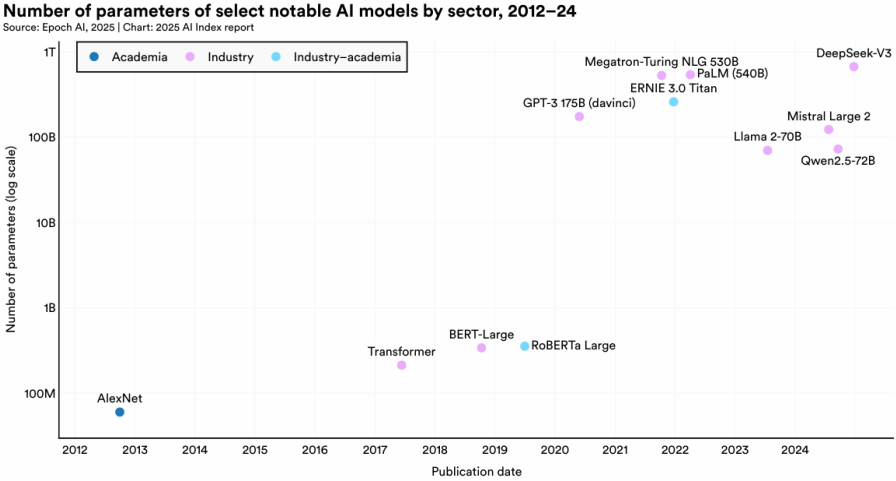


2015 VGG19 (Oxford): Convolutional, 1.45 10^8

2020 GPT-3 (OpenAI): Transformer, 1.75 10^{11}

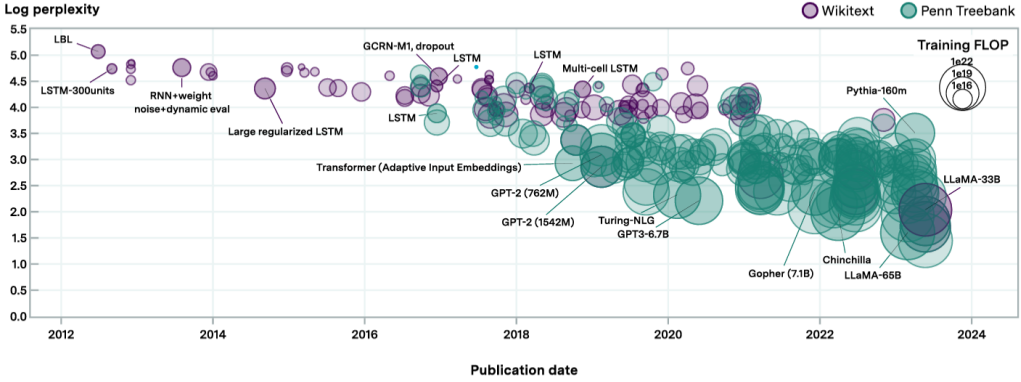
Bernstein, Liane, et al. "Freely scalable and reconfigurable optical hardware for deep learning." Scientific reports 11.1 (2021): 3144.

Deep Learning: Hoy - Cantidad de parámetros



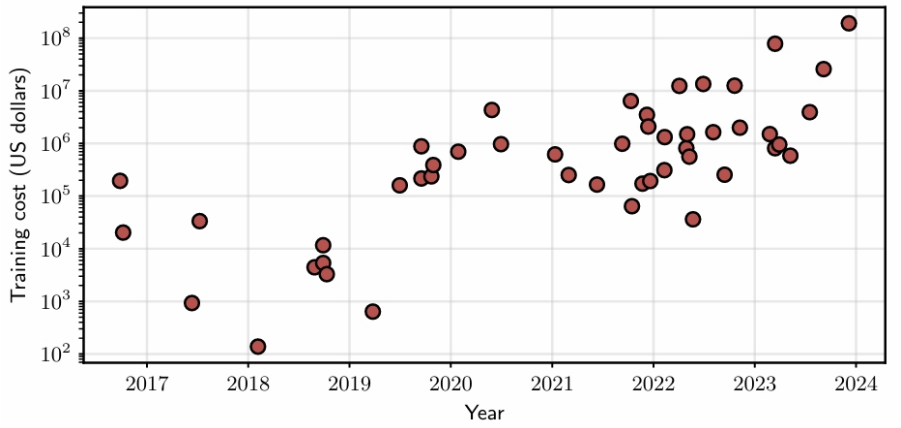
The Stanford 2025 AI Index Report

Deep Learning: Hoy - Performance



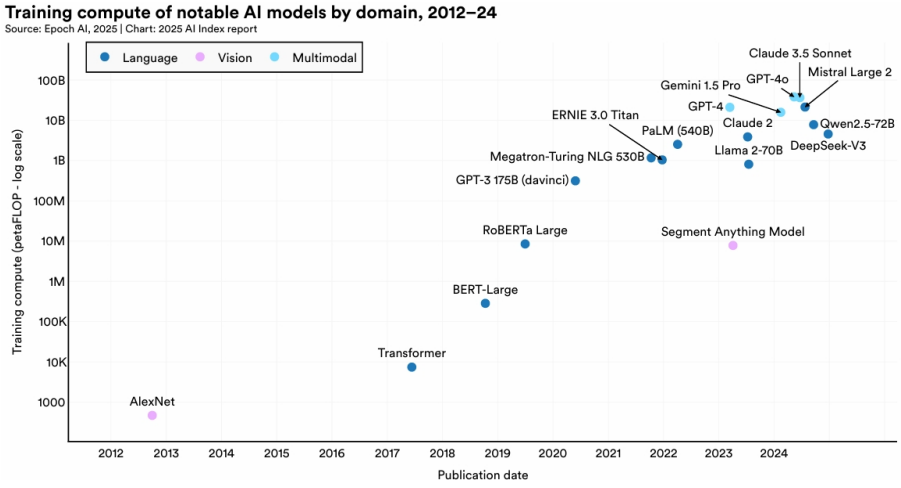
Algorithmic progress in language models 2024 Arxiv

Deep Learning: Hoy - Costo de entrenamiento



The Stanford 2024 AI Index Report

Deep Learning: Hoy - Cómputo de entrenamiento



The Stanford 2025 AI Index Report

Bibliografía

- Deep Learning. Ian Goodfellow, Yoshua Bengio, Aaron Courville. MIT Press, 2016.
- Prince, Simon JD. Understanding deep learning. MIT press, 2023.
- Bishop, Christopher M., and Hugh Bishop. Deep learning: Foundations and concepts. Springer Nature, 2023.
- Scardapane, Simone. "Alice's Adventures in a Differentiable Wonderland–Volume I, A Tour of the Land." arXiv preprint arXiv:2404.17625 (2024).