Introducción a Deep Learning y Series de tiempo

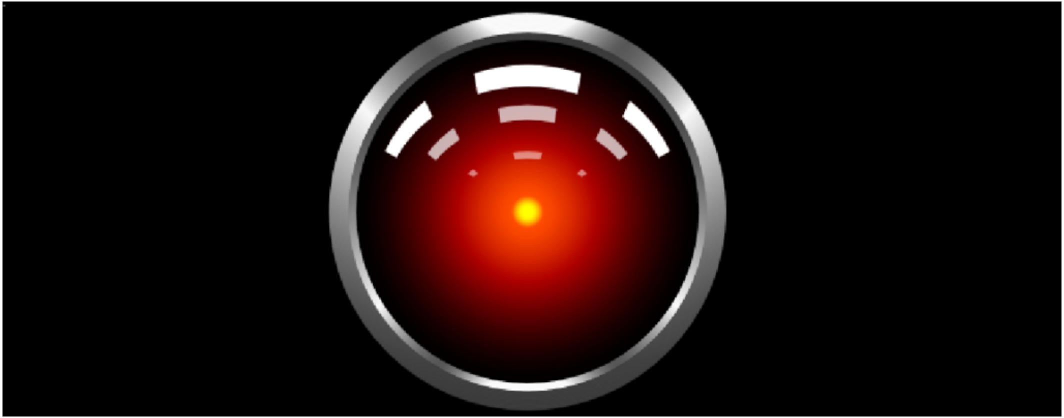
Aprendizaje de Máquina aplicado

Marco Teran

EAFIT

2025

Contenido



▸ ver video

Objetivos y mapa

Objetivos de aprendizaje

- Distinguir IA, ML y Deep Learning (DL).
- Entender tareas: clasificación, regresión, *clustering*.
- Comprender paradigmas: supervisado, no supervisado, aprendizaje por refuerzo (RL).
- Dominar el ciclo CRISP-DM y nociones de MLOps.
- Evaluar con métricas y validación rigurosa.
- Reconocer sesgo–varianza y *overfitting*.

Idea central

ML invierte el paradigma: de reglas explícitas a reglas aprendidas.

Mapa del capítulo

- Historia breve y motivaciones
- Definición formal de ML (Mitchell)
- Taxonomía de tareas y paradigmas
- Ciclo de vida y métricas
- Límites: sesgo–varianza, No Free Lunch
- Aplicaciones y buenas prácticas
- Síntesis y referencias clave

$$\text{ML}: \; f^* = \arg\min_f \mathcal{L}(f(X), y)$$

Un poco de historia...

The thinking machine

ver video

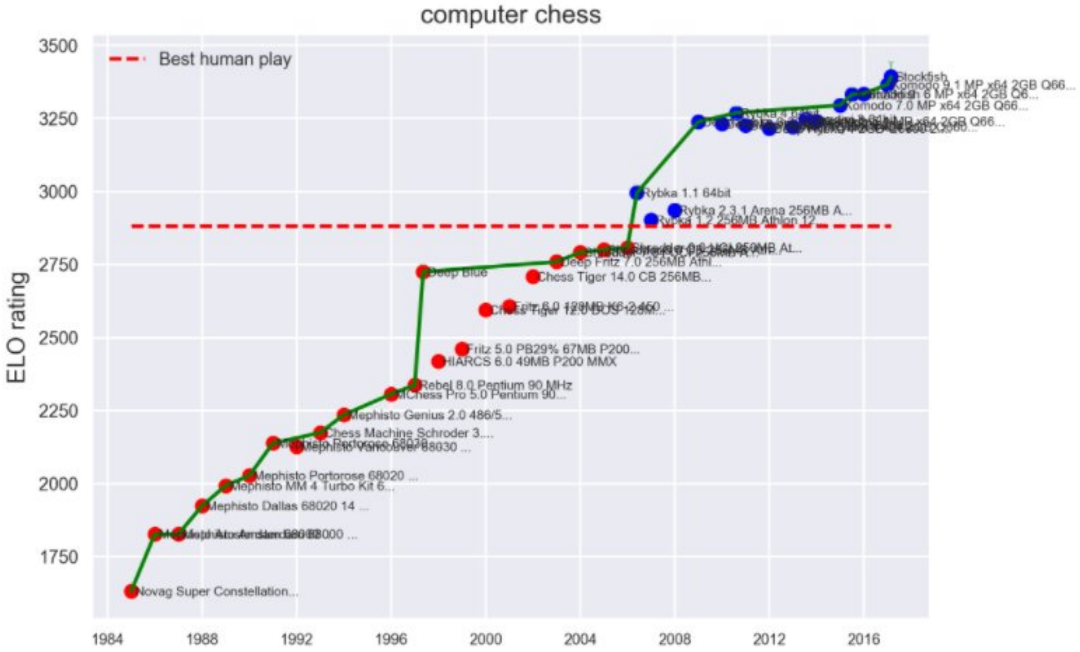
original EN

DeepBlue vs Gasparov (1997)

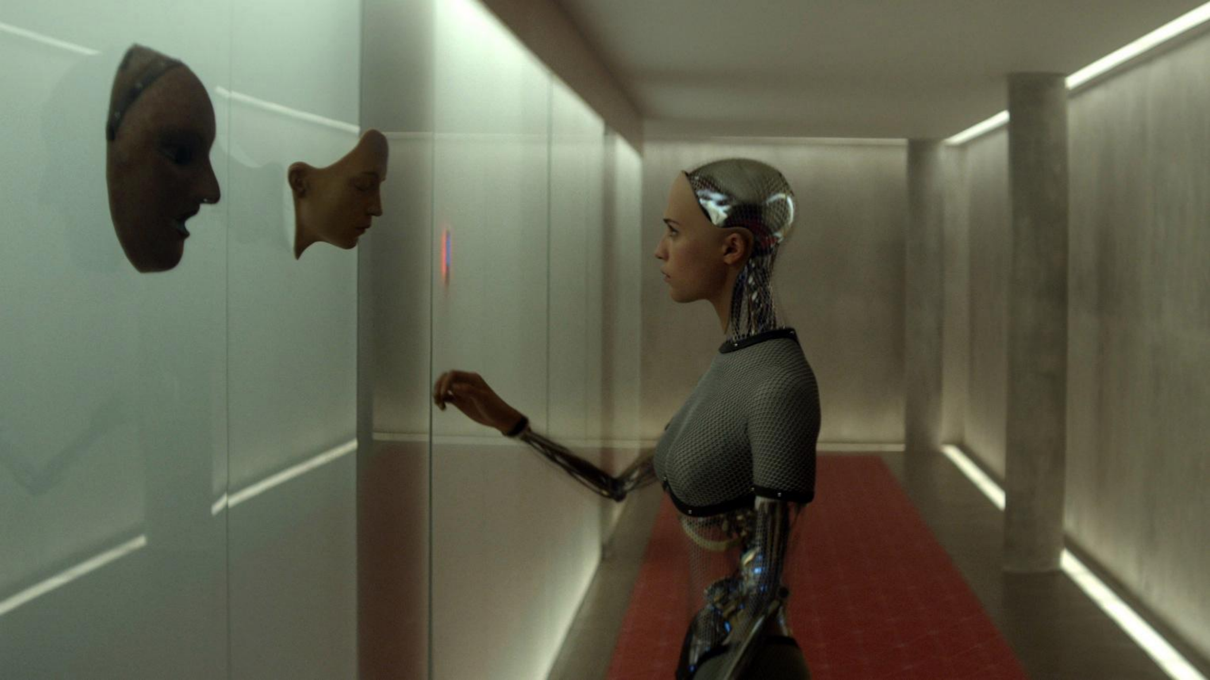
ver video

DeepBlue vs Gasparov (1997)

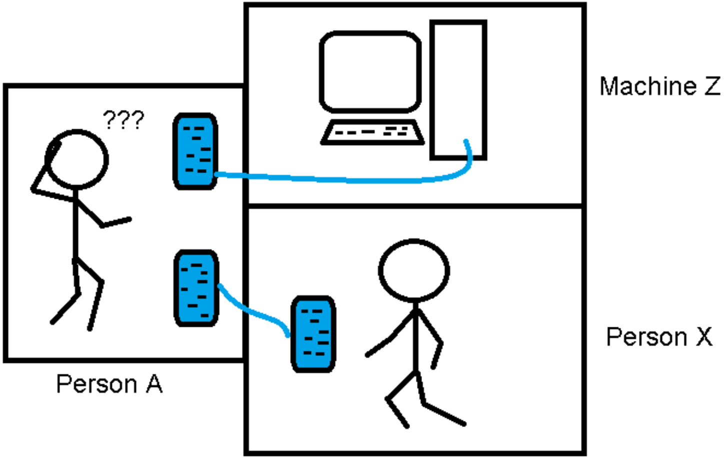
‣ ver video


computer chess
Best human play
ELO rating
3500
3250
3000
2750
2500
2250
2000
1750
1984
1988
1992
1996
2000
2004
2008
2012
2016
Stockfish
Komodo 9.1 MP x64 2GB Q66...
Komodo 7.0 MP x64 2GB Q66...
Rybka 1.1 64bit
Rybka 2.3.1 Arena 256MB A...
Deep Fritz 7.0 256MB Athl...
Deep Blue
Chess Tiger 14.0 CB 256MB...
Fritz 5.0 PB29% 67MB P200...
HIARCS 6.0 49MB P200 MMX
Rebel 8.0 Pentium 90 MHz
Mephisto Genius 2.0 486/5...
Chess Machine Schroder 3....
Mephisto Portorose 68020 ...
Mephisto MM 4 Turbo Kit 6...
Mephisto Dallas 68020 14 ...
Novag Super Constellation...

¿Qué es la Inteligencia Artificial?



El test de Turing



El test de Turing


Hi, how can I help?
© Copyright Webtilia
Webtilia

Google Duplex (2018)

▸ ver video

Inteligencia y racionalidad computacional

- **Inteligencia:** decisiones adecuadas vs. criterio dado
- Requiere conocimiento operacional utilizable
- Racionalidad: maximizar una utilidad definida

$$\text{Agente racional}: \pi^* = \arg\max_{\pi} \mathbb{E}[U \mid \pi]$$

Conexión

De teoría de decisión a agentes: utilidad, riesgo y costo.

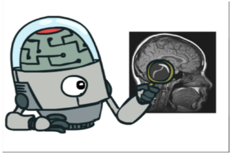
Inteligencia Artificial

La noción de **inteligencia** puede ser definida de varias formas:

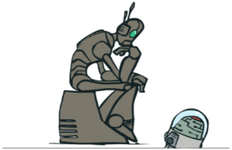
"the ability to take the right decisions, according to some criterion (e.g. survival and reproduction, for most animals)"

La toma de buenas decisiones requiere **conocimiento** en forma **operacional**.

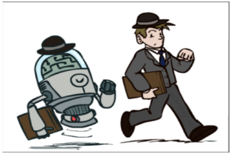
Cuatro enfoques clásicos de IA



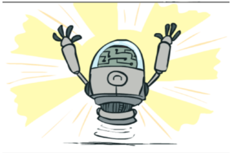
(a) Pensar como humano



(b) Pensar racionalmente



(c) Actuar como humano



(d) Actuar racionalmente

¿Qué es IA?

La ciencia de crear máquinas que:

- Piensen como personas
- Actúen como personas
- Piensen y actúen racionalmente

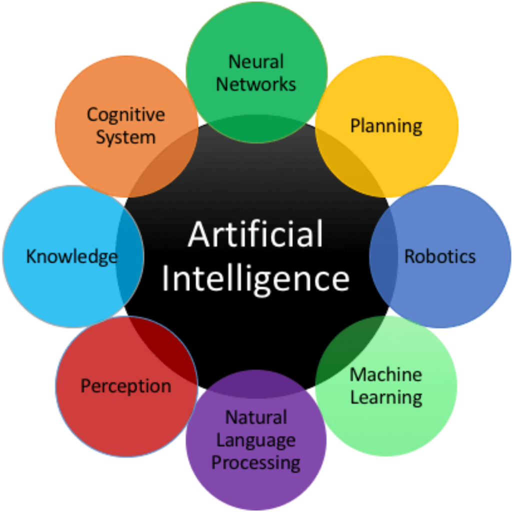
Agente: software inteligente.
Robot: plataforma física de acción.

IA hoy: racionalidad computacional

- Objetivos $→$ función de utilidad
- Agir racional $→$ maximizar utilidad
- Restricciones: tiempo, datos, cómputo

Pitfall

Objetivos mal definidos inducen comportamientos no deseados.



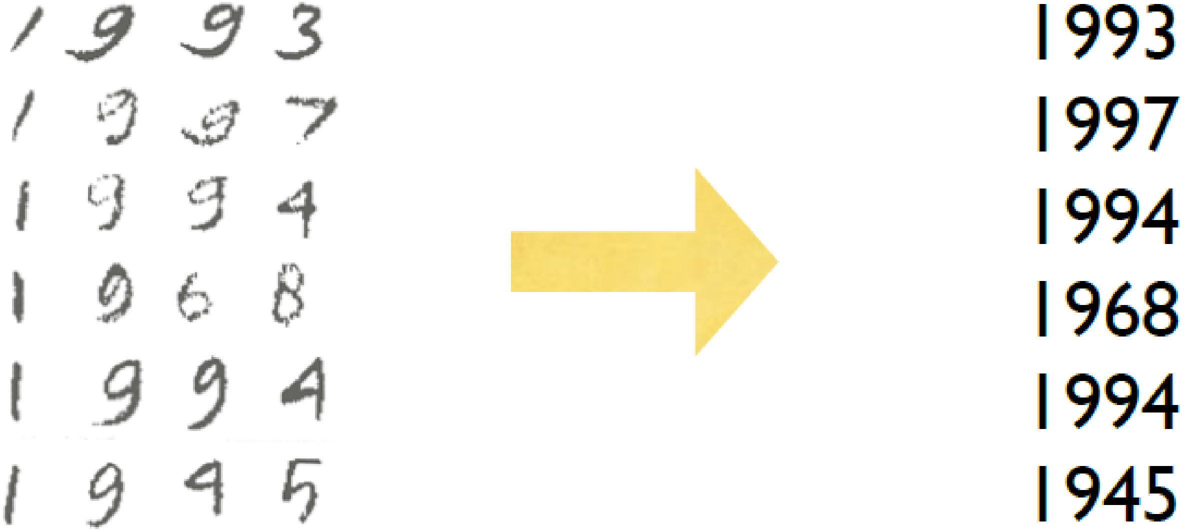
Requerimientos de un agente ideal

- Conocimiento y razonamiento
- Aprendizaje computacional
- Percepción y lenguaje
- Planeación y control
- Robótica e interacción

Aprendizaje Computacional

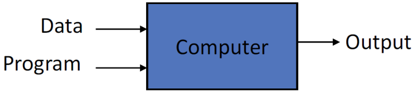
Los computadores hacen lo que les decimos

¿Y si les pedimos aprender y mejorar?

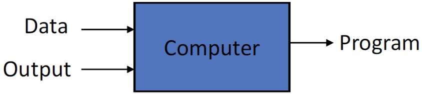

1993
1997
1994
1968
1994
1945

Aprendizaje de máquina vs programación tradicional

Programación tradicional



Machine Learning



$$\text{Reglas (código)} \Rightarrow \text{salidas} \quad \text{vs} \quad \text{Datos} \Rightarrow \text{modelo } f$$

Aprendizaje de máquina: definición operativa

"The acquisition of knowledge or skills through study or experience"

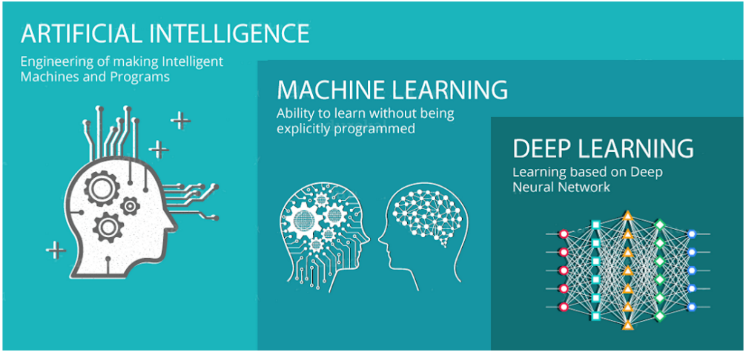
El **machine learning** estudia algoritmos que permiten a un agente *aprender a decidir/actuar* desde datos.

$$f^* = \arg\min_f \mathcal{L}(f(X), y)$$

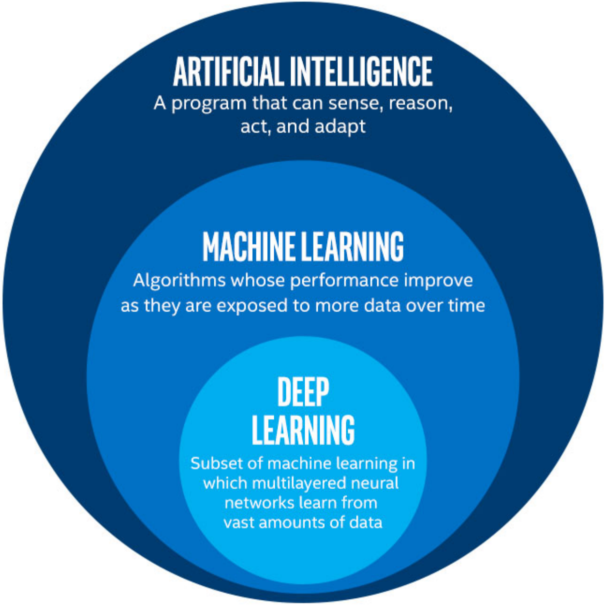
Mitchell (1997)

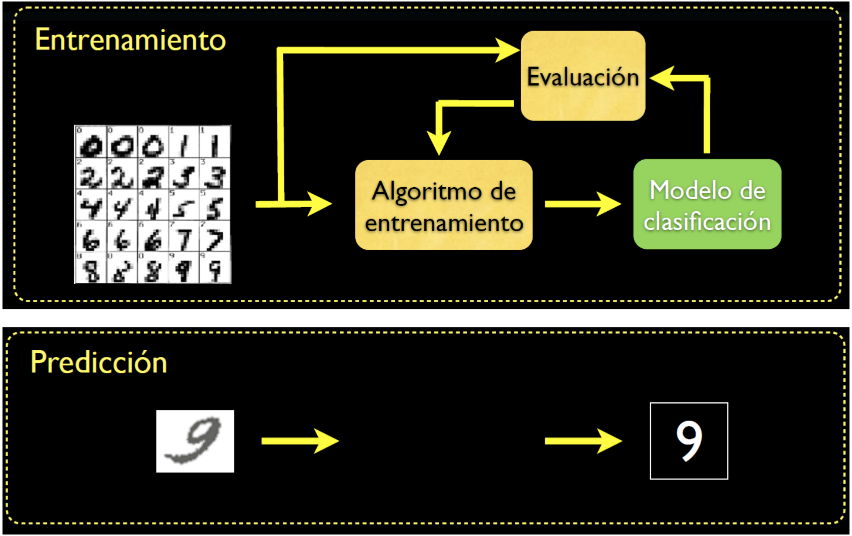
Aprende de experiencia E en tarea T medida por P si P mejora con E.

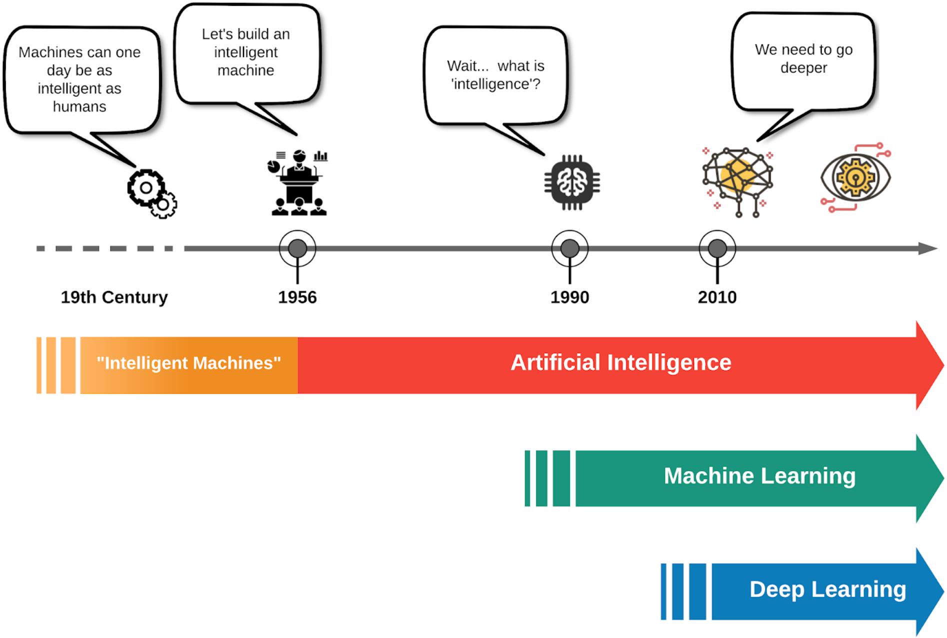
IA, ML y DL: relaciones



- **IA**: imitar comportamientos inteligentes
- **ML**: aprender sin reglas explícitas
- **DL**: patrones con redes neuronales profundas

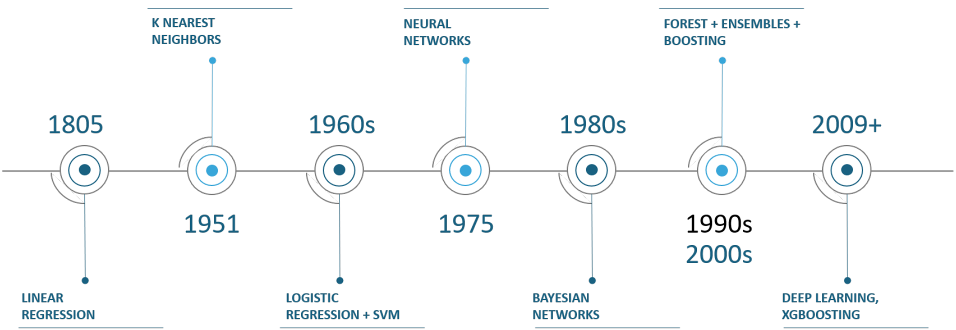

ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data


Entrenamiento
Evaluación
Algoritmo de entrenamiento
Modelo de clasificación
Predicción
9



Machines can one day be as intelligent as humans
Let's build an intelligent machine
Wait... what is 'intelligence'?
We need to go deeper
19th Century
1956
1990
2010
"Intelligent Machines"
Artificial Intelligence
Machine Learning
Deep Learning

Línea del tiempo del Machine Learning



Taxonomía de tareas

- Clasificación: etiquetas discretas
- Regresión: valores continuos
- *Clustering*: estructura sin etiquetas
- Detección de anomalías: casos raros
- Ranking y recomendación: orden y preferencia

$$\text{Supervisado}: \ (x, y) \qquad \text{No supervisado}: \ x$$

Paradigmas de aprendizaje

- Supervisado: pares (x, y) etiquetados
- No supervisado: modela $p(x)$
- Refuerzo (RL): política π con recompensas

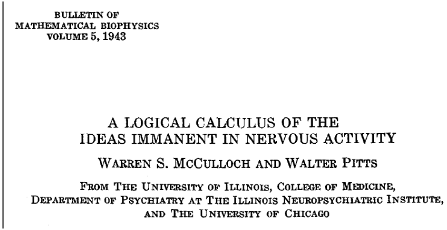
$$\text{RL}: \; s_t \xrightarrow{a_t} r_t, s_{t+1}, \quad \pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_t \gamma^t r_t\right]$$

Conexión

Auto-supervisado: etiquetas generadas desde los propios datos.

Redes Neuronales

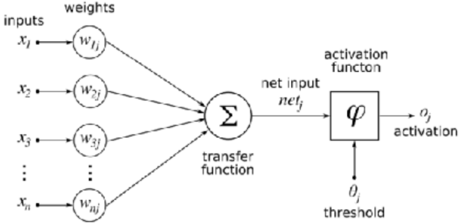
McCulloch & Pitts Artificial Neuron

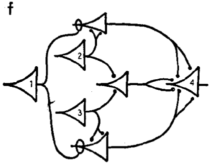

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

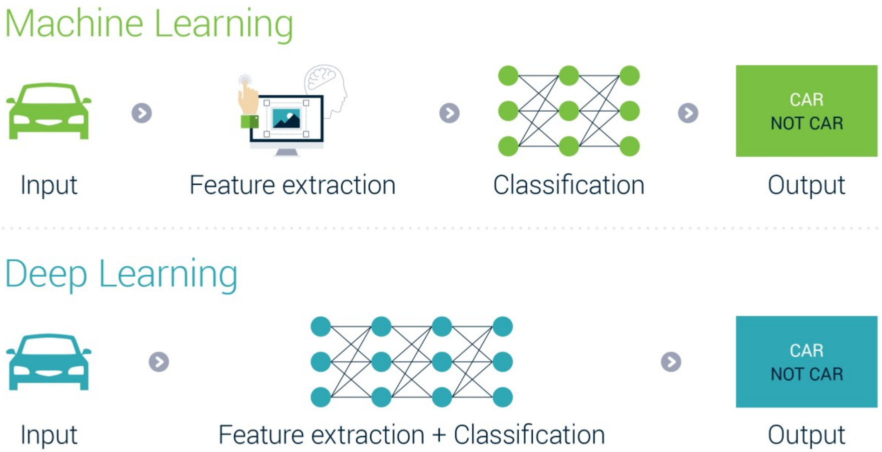
A LOGICAL CALCULUS OF THE
IDEAS IMMANENT IN NERVOUS ACTIVITY

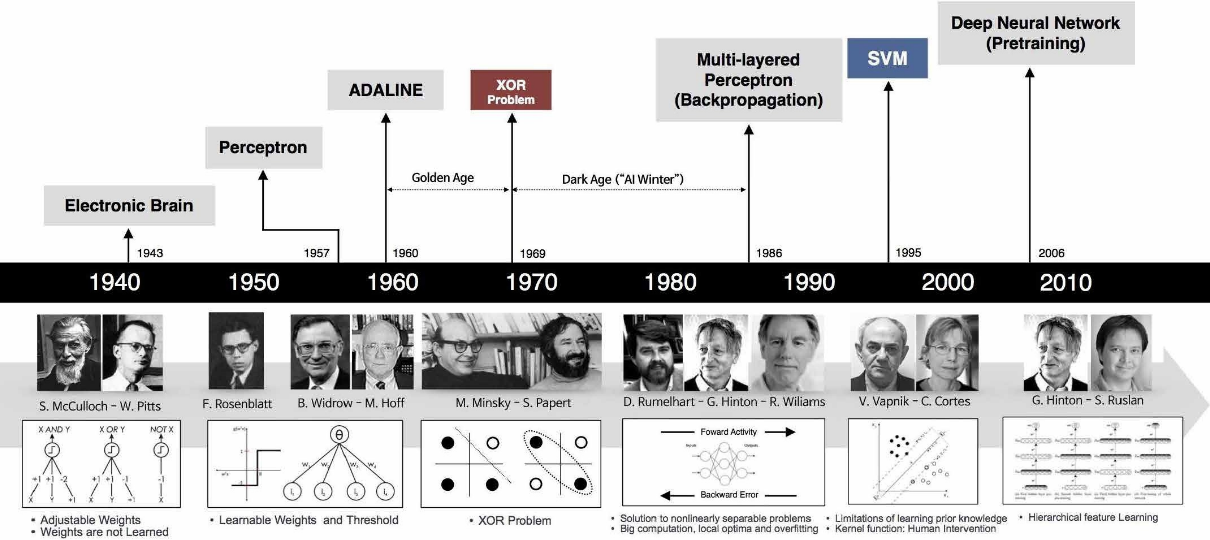
WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

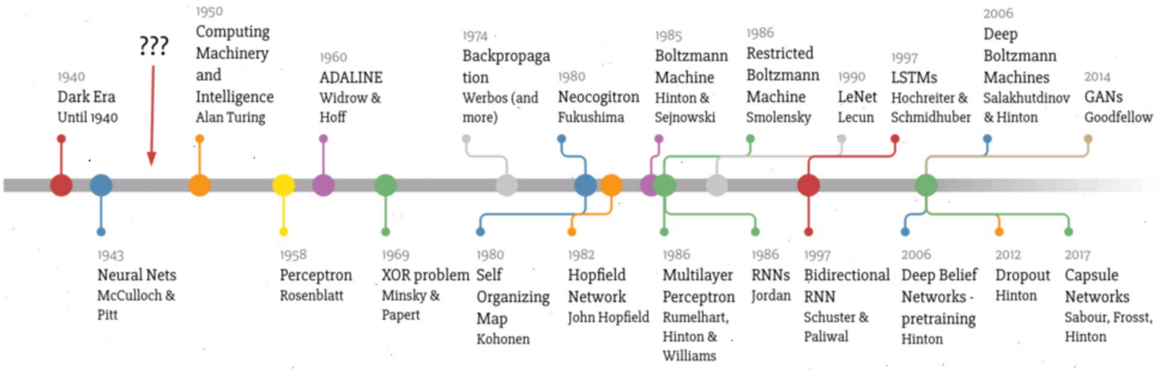




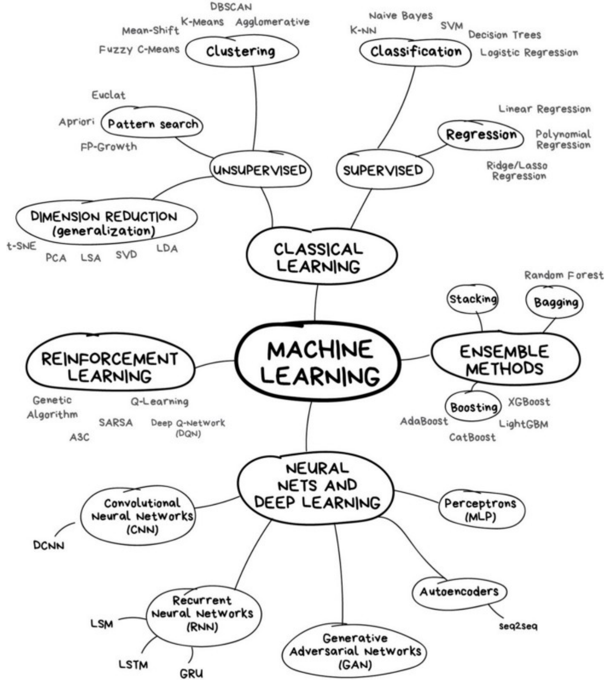

Machine Learning
CAR
NOT CAR
Input
Feature extraction
Classification
Output
Deep Learning
CAR
NOT CAR
Input
Feature extraction + Classification
Output

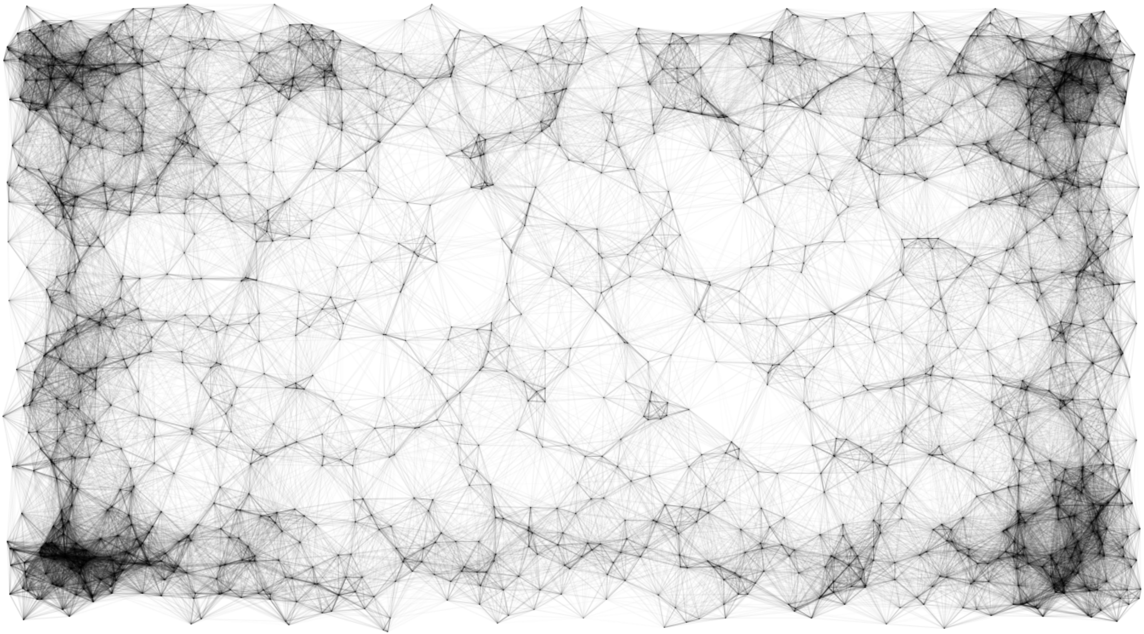

Electronic Brain
Perceptron
ADALINE
XOR
Problem
Multi-layered
Perceptron
(Backpropagation)
SVM
Deep Neural Network
(Pretraining)
Golden Age
Dark Age ("AI Winter")
1943
1957
1960
1969
1986
1995
2006
1940
1950
1960
1970
1980
1990
2000
2010
S. McCulloch - W. Pitts
F. Rosenblatt
B. Widrow - M. Hoff
M. Minsky - S. Papert
D. Rumelhart - G. Hinton - R. Wiliams
V. Vapnik - C. Cortes
G. Hinton - S. Ruslan
X AND Y
X OR Y
NOT X
Foward Activity
Backward Error
• Adjustable Weights
• Weights are not Learned
• Learnable Weights and Threshold
• XOR Problem
• Solution to nonlinearly separable problems
• Big computation, local optima and overfitting
• Limitations of learning prior knowledge
• Kernel function: Human Intervention
• Hierarchical feature Learning

Deep Learning Timeline



Made by Favio Vázquez

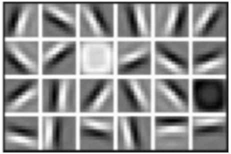

MACHINE LEARNING
CLASSICAL LEARNING
UNSUPERVISED
SUPERVISED
Clustering
DBSCAN
K-Means
Agglomerative
Mean-Shift
Fuzzy C-Means
Classification
Naive Bayes
K-NN
SVM
Decision Trees
Logistic Regression
Pattern search
Euclat
Apriori
FP-Growth
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
DIMENSION REDUCTION (generalization)
t-SNE
PCA
LSA
SVD
LDA
ENSEMBLE METHODS
Stacking
Bagging
Random Forest
Boosting
XGBoost
AdaBoost
LightGBM
CatBoost
REINFORCEMENT LEARNING
Genetic Algorithm
Q-Learning
SARSA
Deep Q-Network (DQN)
A3C
NEURAL NETS AND DEEP LEARNING
Convolutional Neural Networks (CNN)
DCNN
Perceptrons (MLP)
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq



¿Por qué Deep Learning y por qué ahora?

¿Por qué el Deep Learning?

- Las características diseñadas a mano consumen mucho tiempo, son frágiles y no se pueden escalar en la práctica.
- ¿Podemos aprender las **características subyacentes** directamente de los datos?

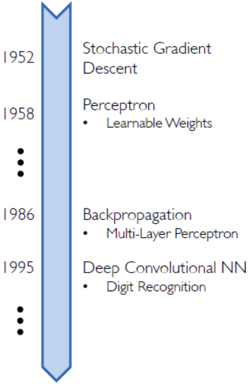


(g) Características de bajo nivel: Líneas y bordes

(h) Características de nivel medio: Ojos, nariz y oídos

(i) Características de alto nivel: Estructura facial

¿Por qué ahora



Las redes neuronales se remontan a décadas atrás, así que ¿por qué el resurgimiento?

1. Big Data

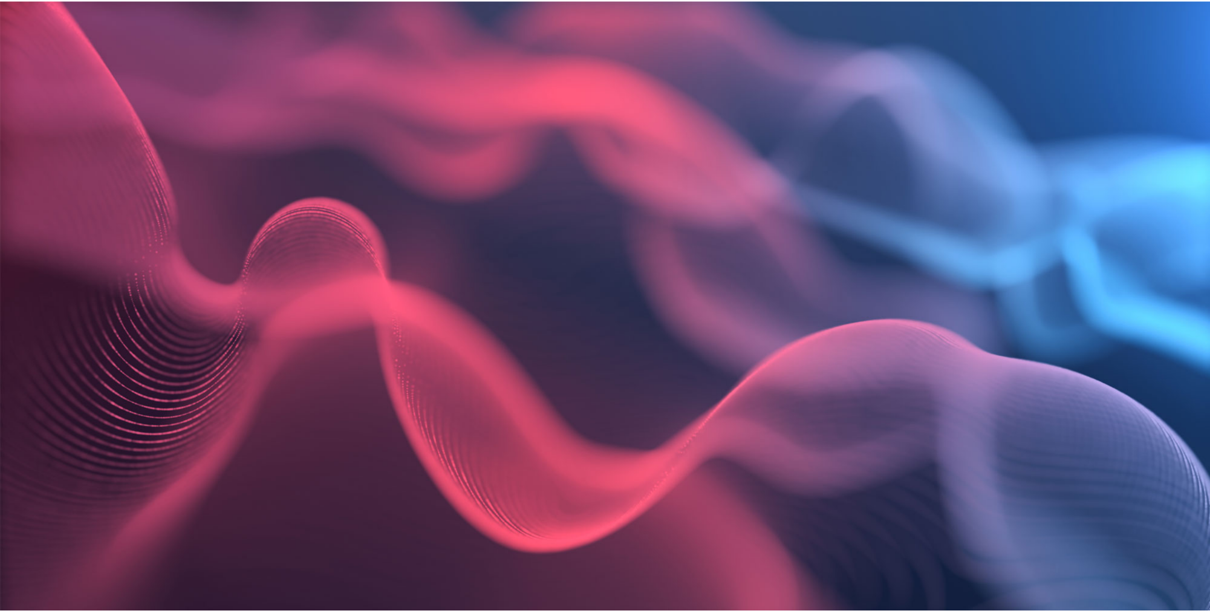
- Conjuntos de datos más grandes
- Recolección y almacenamiento más fácil

2. Hardware

- Unidades de procesamiento gráfico (GPU)
- Masivamente paralelizable

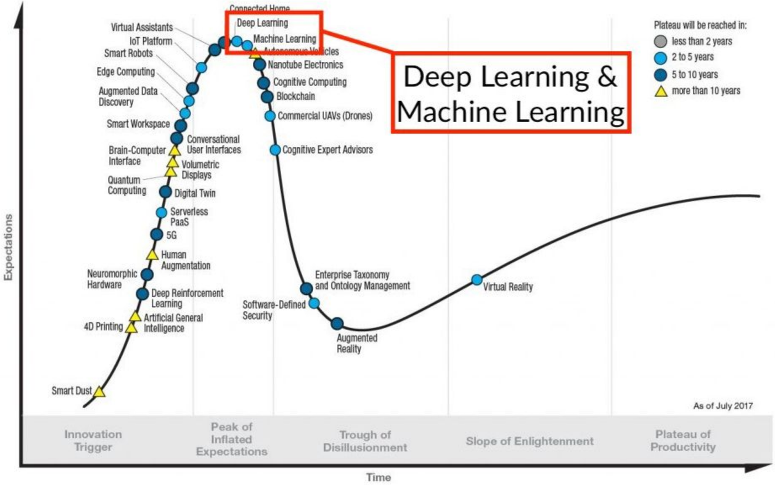
3. Software

- Técnicas mejoradas
- Nuevos modelos
- Toolboxes



Aplicaciones

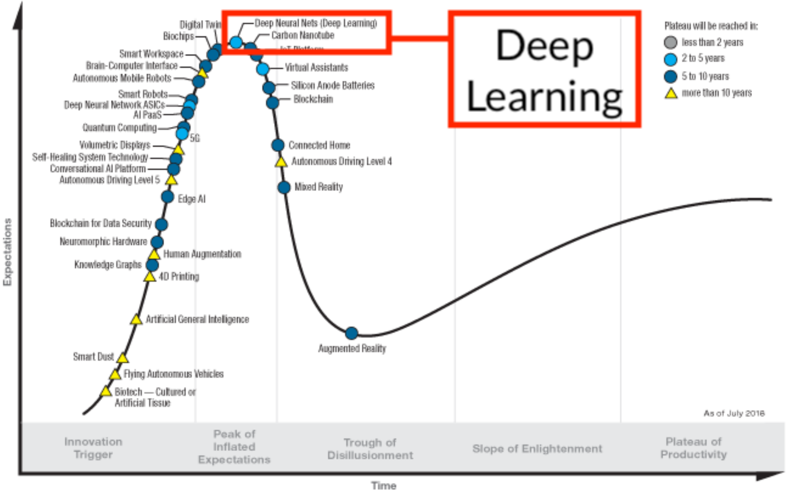
Gartner Hype Cycle for Emerging Technologies, 2017



gartner.com/SmarterWithGartner

Source: Gartner (July 2017)

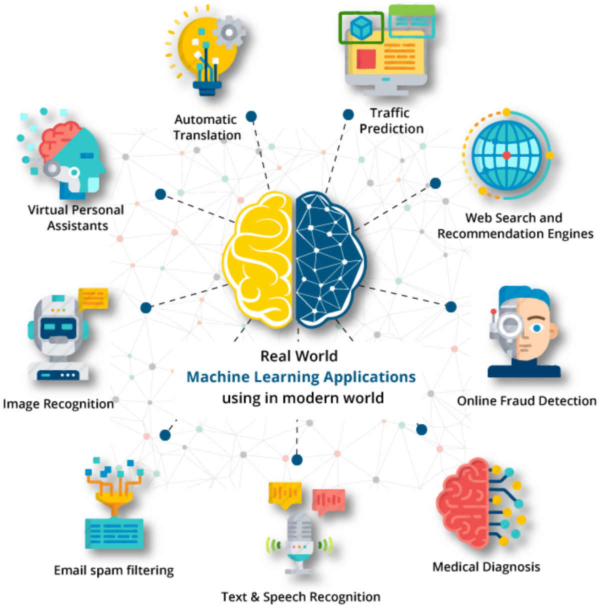
Hype Cycle for Emerging Technologies, 2018

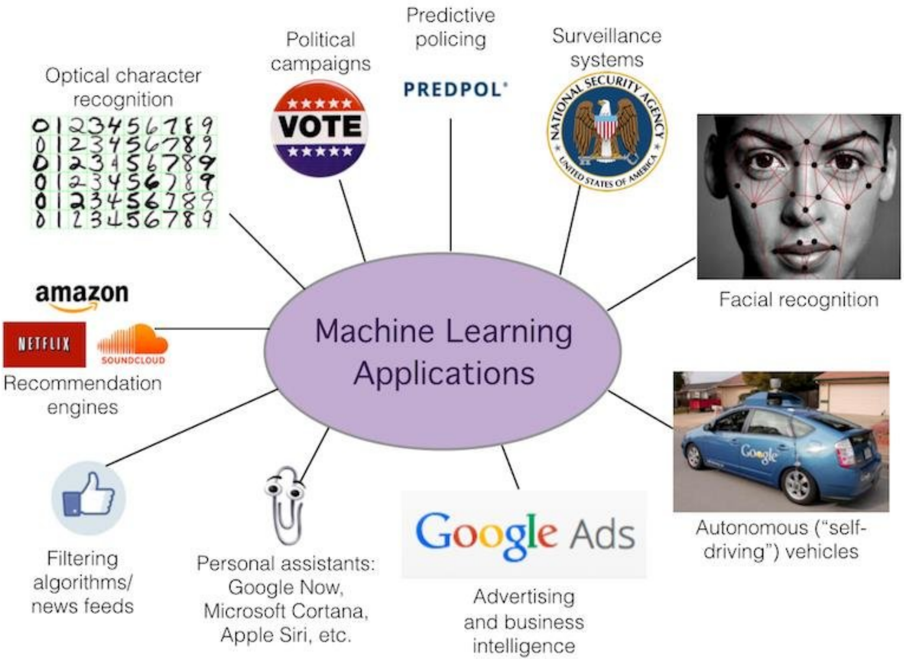


gartner.com/SmarterWithGartner

Source: Gartner (August 2018)

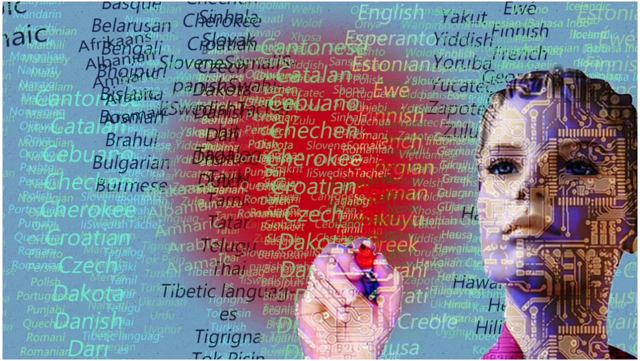
Gartner.


Automatic Translation
Traffic Prediction
Virtual Personal Assistants
Web Search and Recommendation Engines
Real World
Machine Learning Applications
using in modern world
Image Recognition
Online Fraud Detection
Email spam filtering
Text & Speech Recognition
Medical Diagnosis

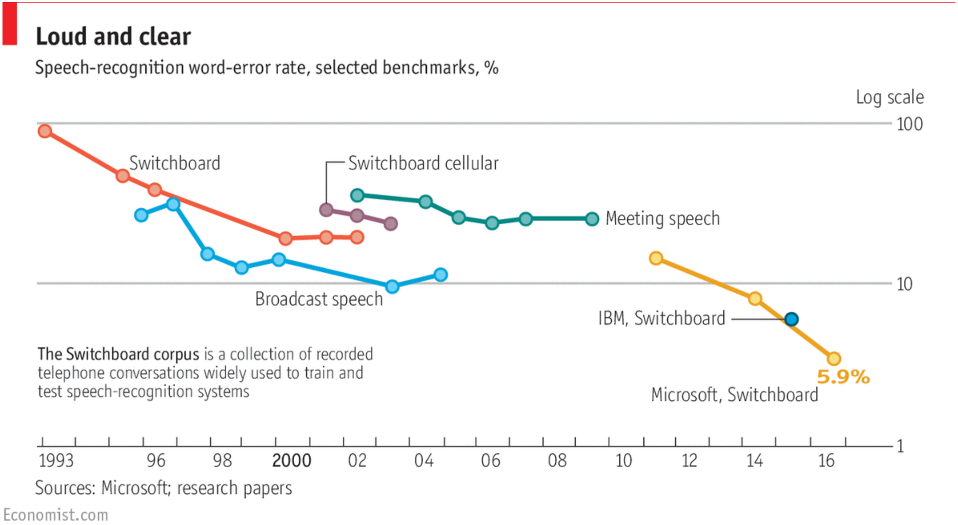

Machine Learning Applications
Optical character recognition
0123456789
Political campaigns
VOTE
Predictive policing
PREDPOL
Surveillance systems
NATIONAL SECURITY AGENCY
UNITED STATES OF AMERICA
Facial recognition
amazon
NETFLIX
SOUNDCLOUD
Recommendation engines
Autonomous ("self-driving") vehicles
Google
Filtering algorithms/ news feeds
Personal assistants: Google Now, Microsoft Cortana, Apple Siri, etc.
Google Ads
Advertising and business intelligence

Procesamiento de lenguaje natural

- Dado un texto, predecir la temática
- Dado un email, predecir si este es un spam
- Dado un texto, predecir el idioma y su traducción a otro lenguaje



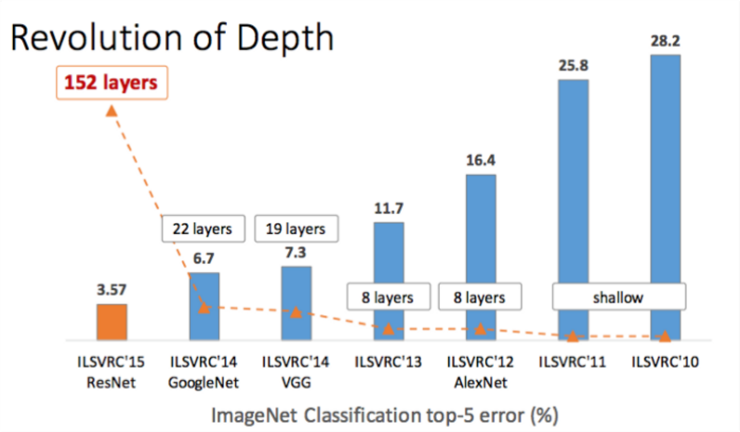
Reconocimiento de Voz



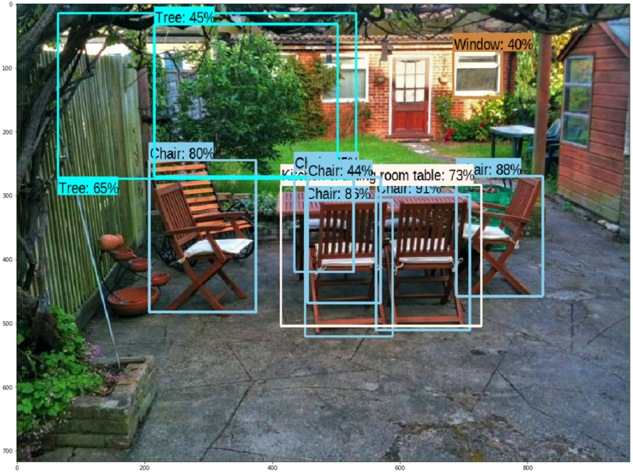
Reconocimiento de Voz

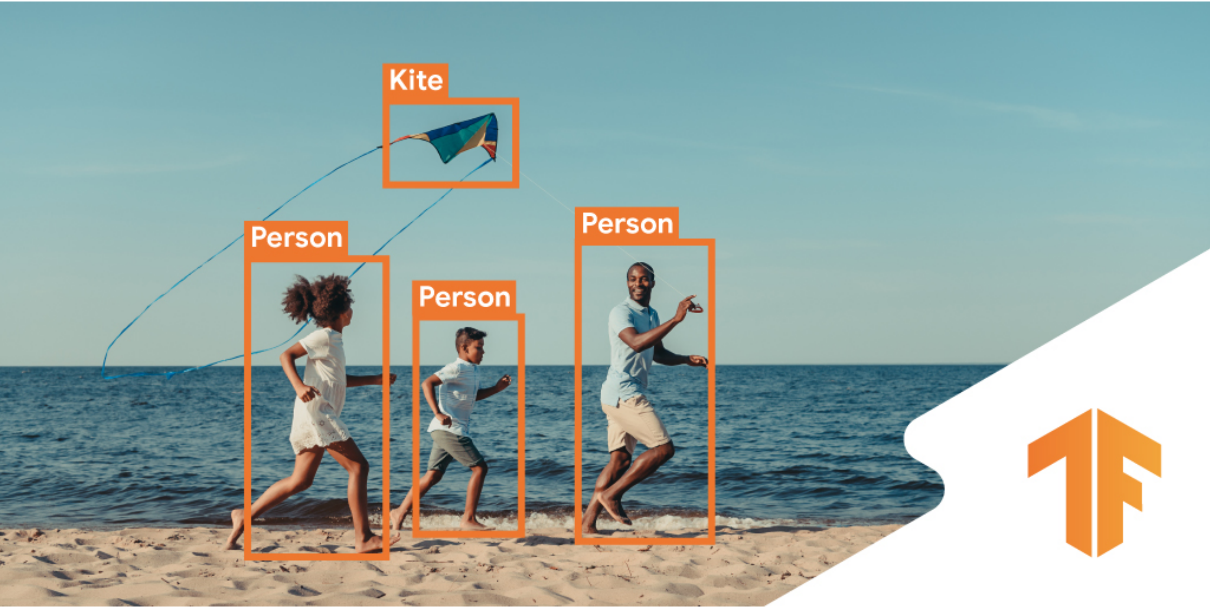
ver video

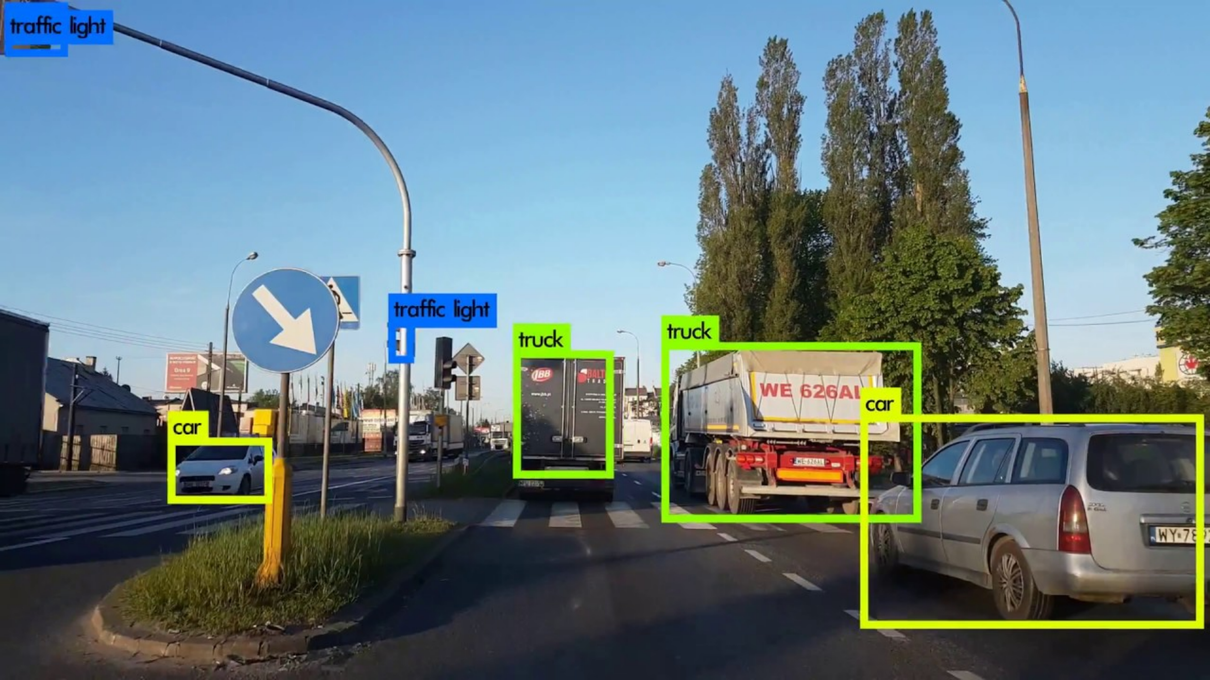
Visión por Computador



Detección de objetos

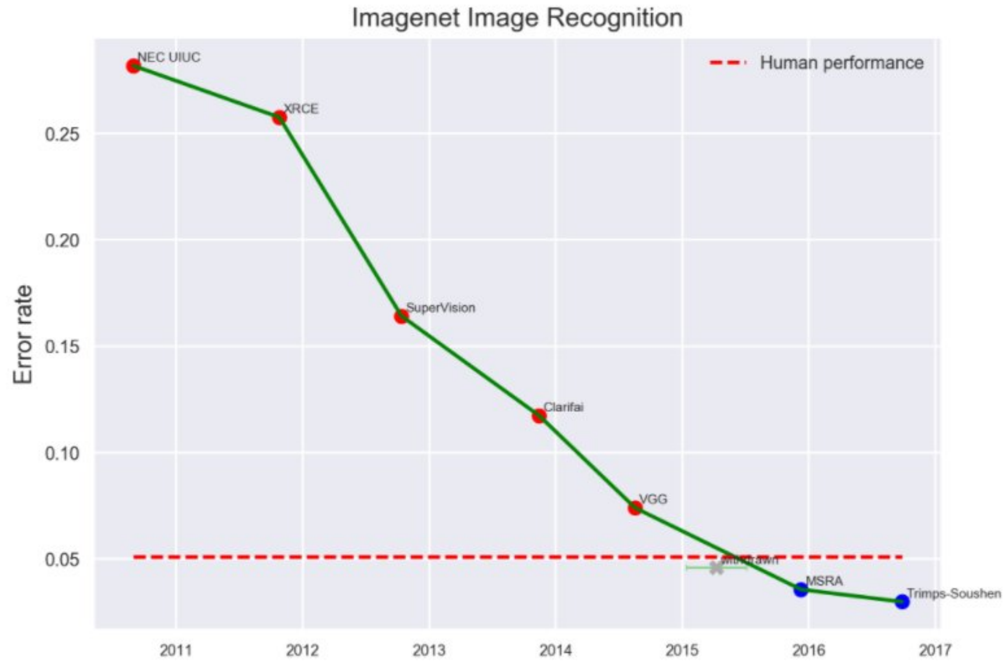



Kite
Person
Person
Person

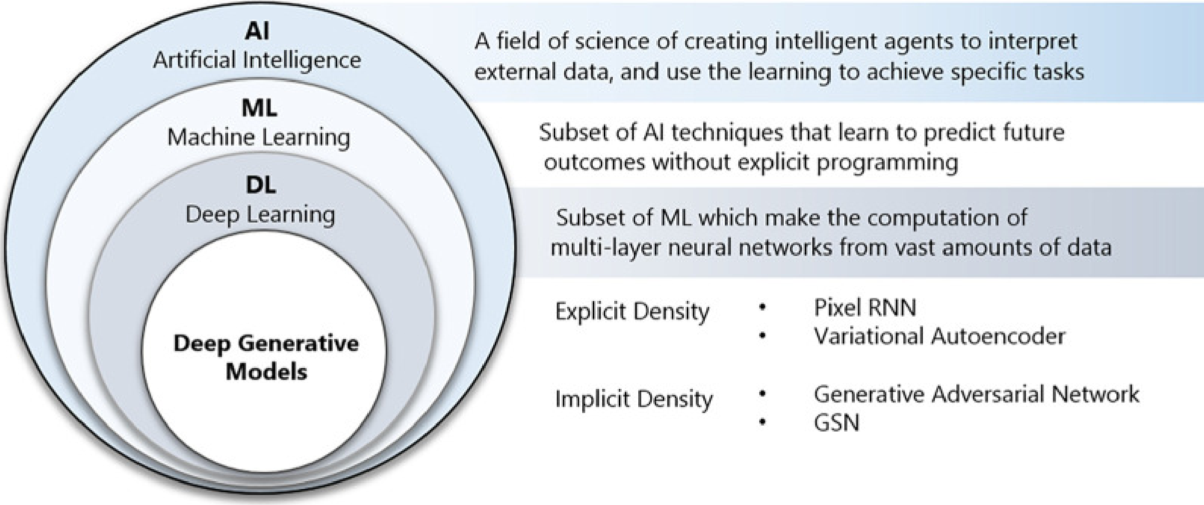

traffic light
traffic light
truck
truck
car
car

Detección de objetos

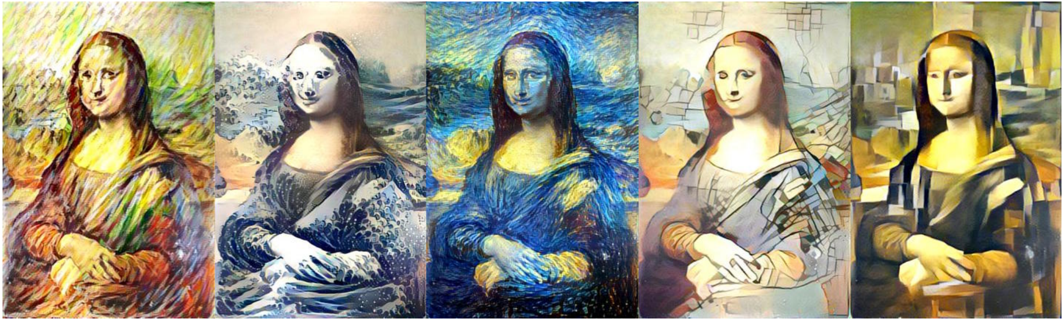
demo


Imagenet Image Recognition
Human performance
Error rate
0.25
0.20
0.15
0.10
0.05
NEC UIUC
XRCE
SuperVision
Clarifai
VGG
MSRA
Trimps-Soushen
2011
2012
2013
2014
2015
2016
2017

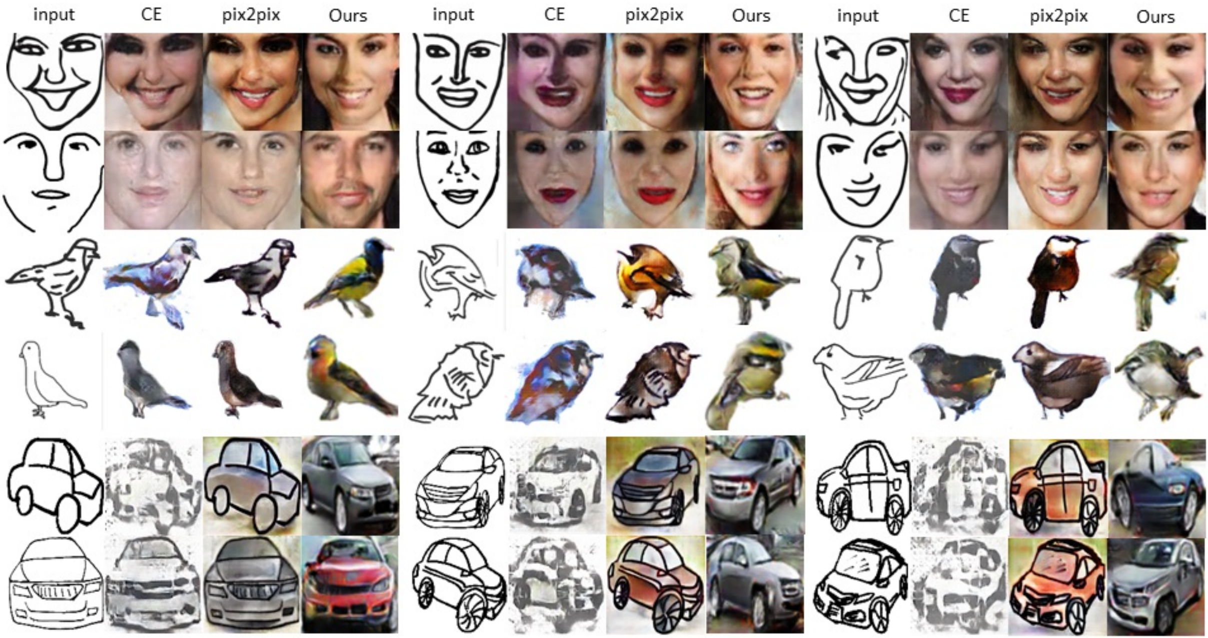
Semantic Segmentation


AI
Artificial Intelligence
ML
Machine Learning
DL
Deep Learning
Deep Generative Models
A field of science of creating intelligent agents to interpret external data, and use the learning to achieve specific tasks
Subset of AI techniques that learn to predict future outcomes without explicit programming
Subset of ML which make the computation of multi-layer neural networks from vast amounts of data
Explicit Density
Pixel RNN
Variational Autoencoder
Implicit Density
Generative Adversarial Network
GSN

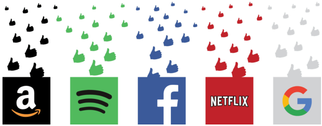
Style transfer

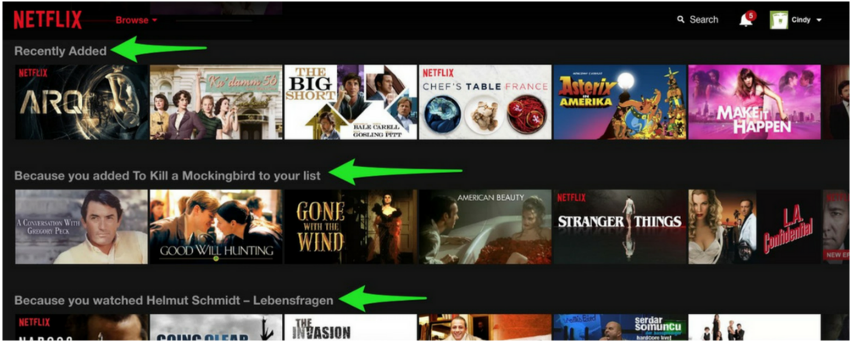


Style transfer


input
CE
pix2pix
Ours
input
CE
pix2pix
Ours
input
CE
pix2pix
Ours

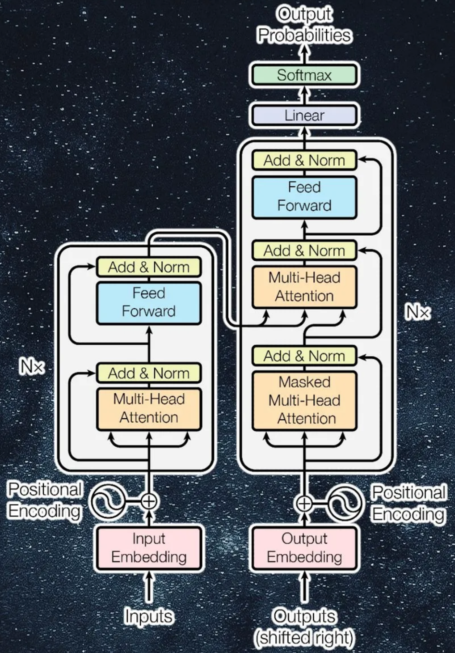
Recommender systems

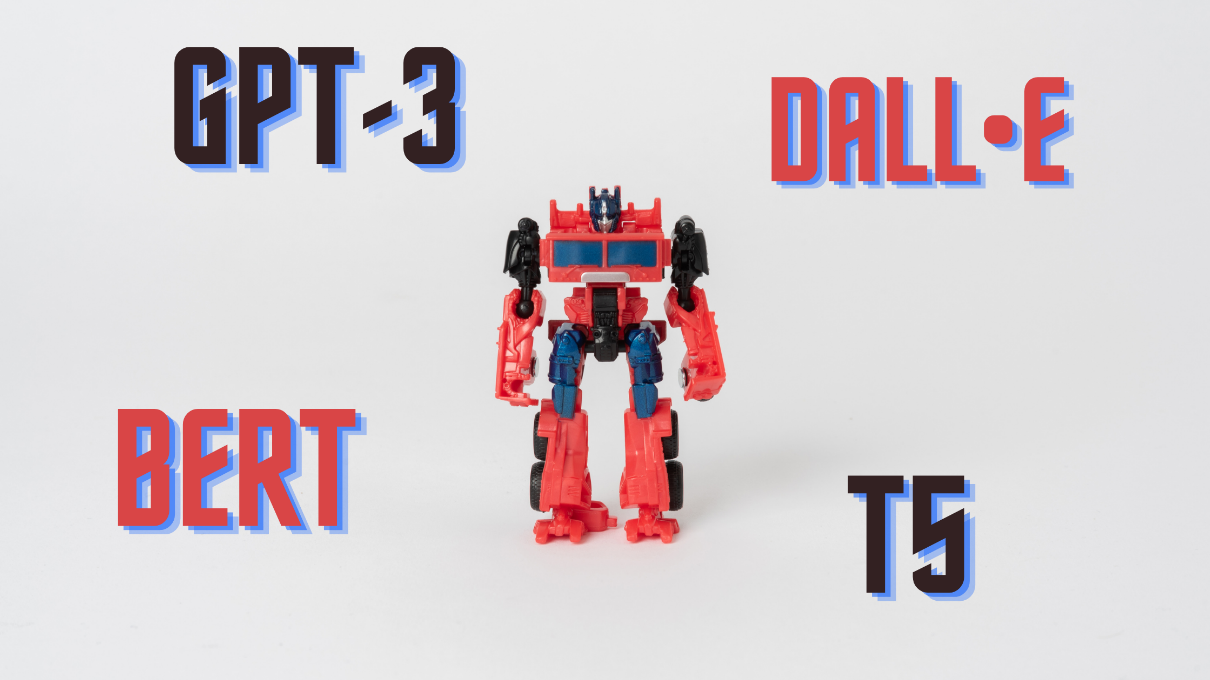




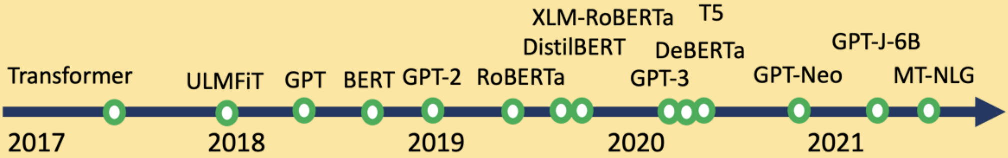

source
destination
Coarse styles copied

Transformers: La nueva joya del aprendizaje profundo


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Input Embedding
Inputs
Positional Encoding
Output Embedding
Outputs (shifted right)


GPT-3
DALL·E
BERT
T5

TRANSFORMERS



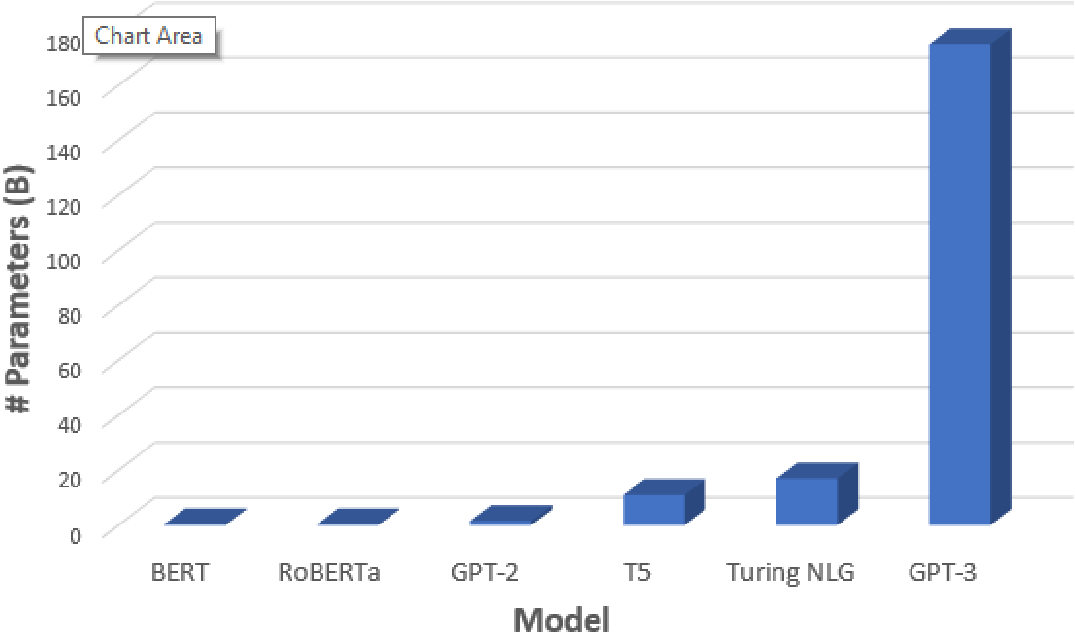
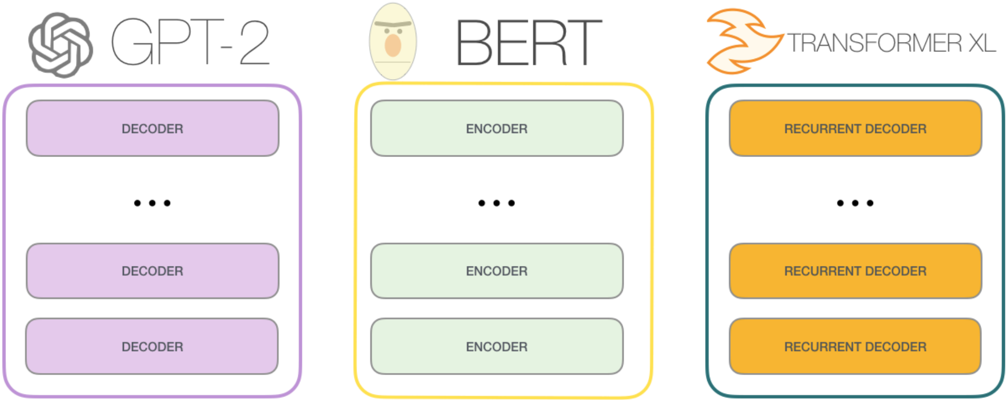
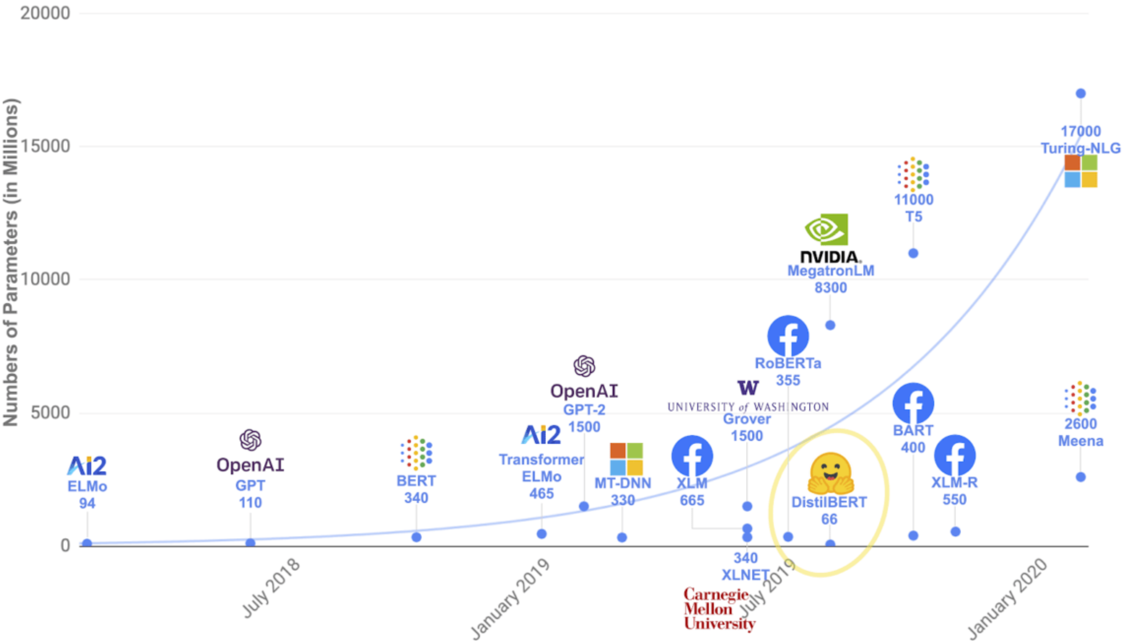
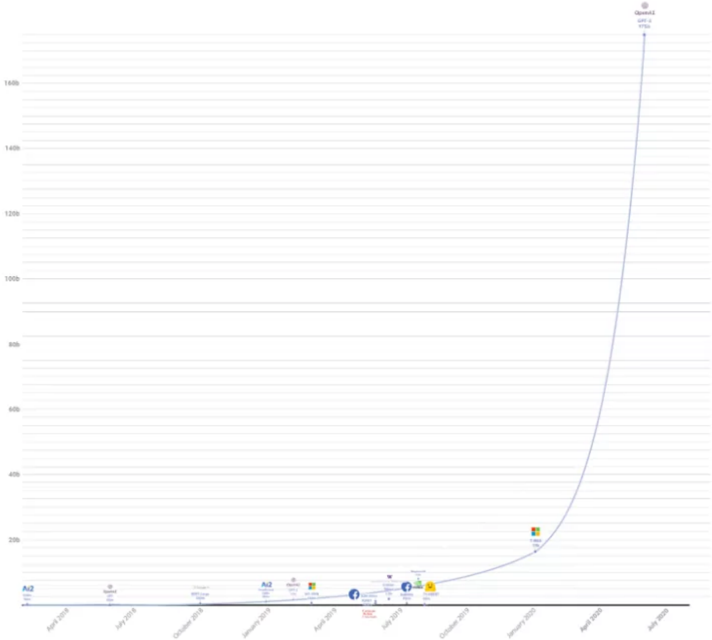


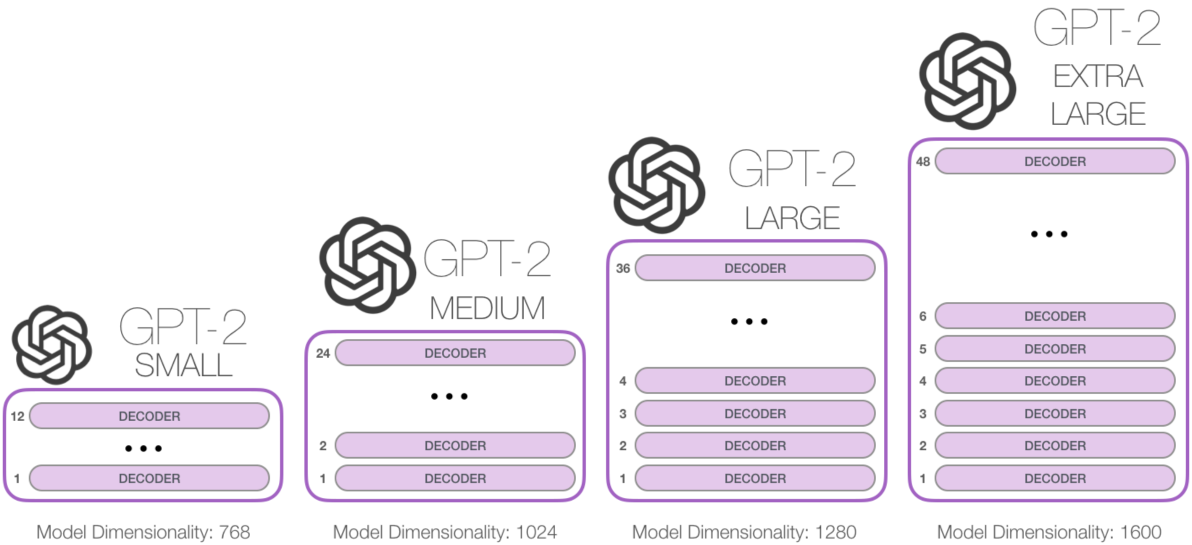
Chart Area
180
160
140
120
100
80
60
40
20
0
Parameters (B)
BERT
RoBERTa
GPT-2
T5
Turing NLG
GPT-3
Model


GPT-2
DECODER
DECODER
DECODER
BERT
ENCODER
ENCODER
ENCODER
TRANSFORMER XL
RECURRENT DECODER
RECURRENT DECODER
RECURRENT DECODER



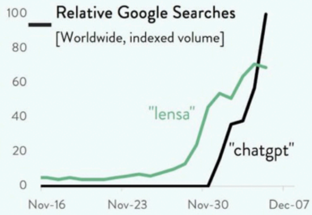
Numbers of Parameters (in Millions)
20000
15000
10000
5000
0
July 2018
January 2019
July 2019
January 2020
ELMo
94
GPT
110
BERT
340
Transformer
ELMo
465
GPT-2
1500
MT-DNN
330
XLM
665
Grover
1500
340
XLNET
RoBERTa
355
DistilBERT
66
MegatronLM
8300
11000
T5
BART
400
XLM-R
550
17000
Turing-NLG
2600
Meena

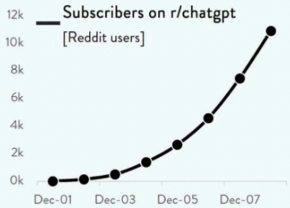

160b
140b
120b
100b
80b
60b
40b
20b
April 2018
July 2018
October 2018
January 2019
April 2019
July 2019
October 2019
January 2020
April 2020
July 2020

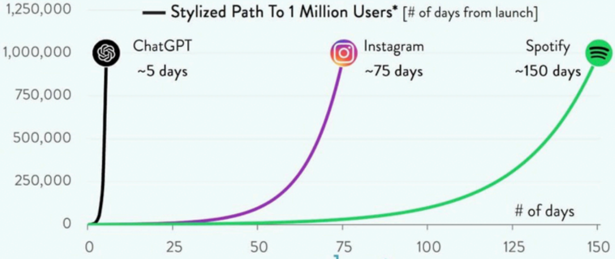

GPT-2 SMALL
12 DECODER
1 DECODER
Model Dimensionality: 768
GPT-2 MEDIUM
24 DECODER
2 DECODER
1 DECODER
Model Dimensionality: 1024
GPT-2 LARGE
36 DECODER
4 DECODER
3 DECODER
2 DECODER
1 DECODER
Model Dimensionality: 1280
GPT-2 EXTRA LARGE
48 DECODER
6 DECODER
5 DECODER
4 DECODER
3 DECODER
2 DECODER
1 DECODER
Model Dimensionality: 1600

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or "GPT-3"	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

ChatGPT From OpenAI Is A Bot Taking The Tech World By Storm







Sources: Google, Subredditstats, Media Reports

chartr

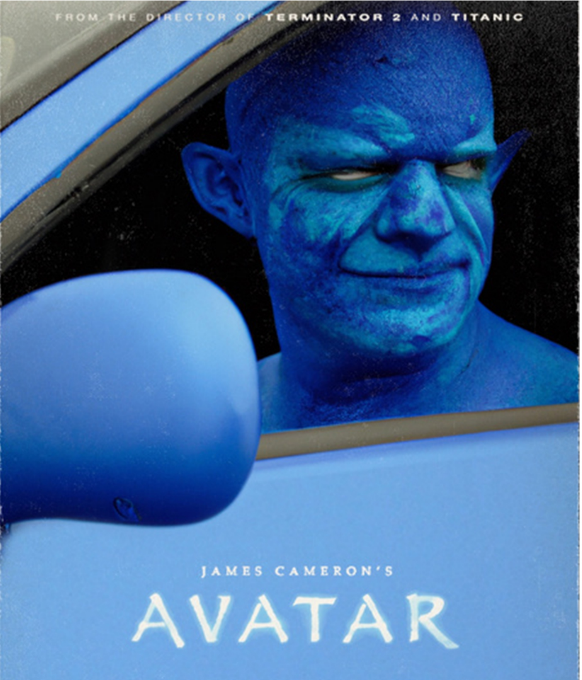
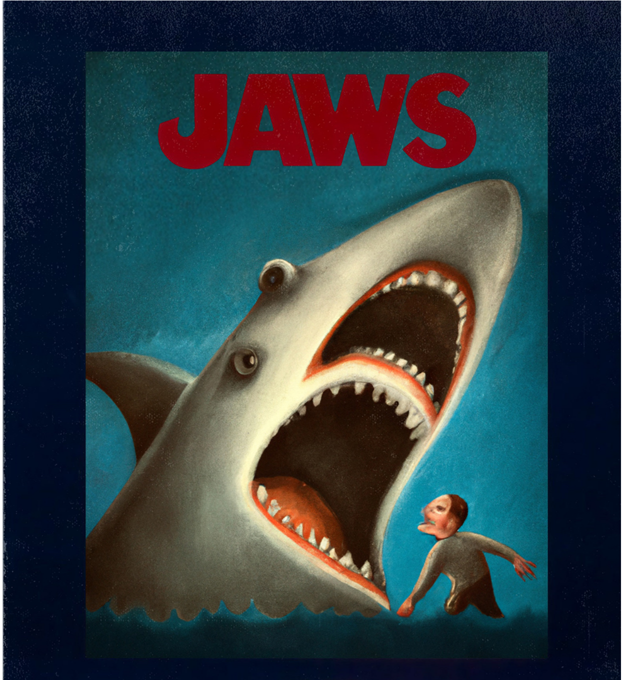
*Path is stylized to the 1m milestone

OpenAI

DALL-E

An Adventure
65 Million Years In The Making.
A STEVEN SPIELBERG FILM
JURASSIC PARK
UNIVERSAL PICTURES PRESENTS AN AMBLIN ENTERTAINMENT PRODUCTION SAM NEILL LAURA DERN JEFF GOLDBLUM
AND RICHARD ATTENBOROUGH "JURASSIC PARK" BOB PECK MARTIN FERRERO B.D. WONG SAMUEL L. JACKSON WAYNE KNIGHT
JOSEPH MAZZELLO ARIANA RICHARDS STAN WINSTON DENNIS MUREN PHIL TIPPETT
MICHAEL LANTIERI JOHN WILLIAMS MICHAEL KAHN RICK CARTER DEAN CUNDEY
MICHAEL CRICHTON MICHAEL CRICHTON AND DAVID KOEPP KATHLEEN KENNEDY AND GERALD R. MOLEN

A STEVEN SPIELBERG FILM
E.T.
THE EXTRA-TERRESTRIAL
IN HIS ADVENTURE ON EARTH
A STEVEN SPIELBERG FILM E.T. THE EXTRA-TERRESTRIAL
DEE WALLACE · PETER COYOTE · HENRY THOMAS AS ELLIOTT · MUSIC BY JOHN WILLIAMS
WRITTEN BY MELISSA MATHISON · PRODUCED BY STEVEN SPIELBERG & KATHLEEN KENNEDY
DIRECTED BY STEVEN SPIELBERG · A UNIVERSAL PICTURE


FROM THE DIRECTOR OF TERMINATOR 2 AND TITANIC
JAMES CAMERON'S
AVATAR

JAWS

DALL·E 1

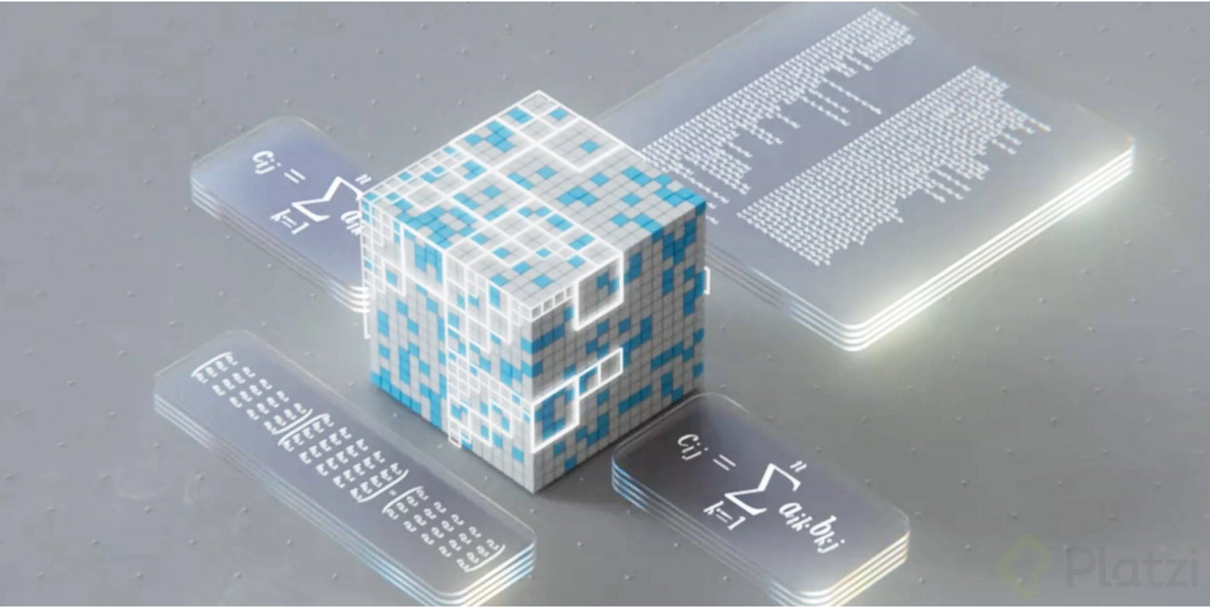
DALL·E 2


670 J. VERMEER
1632-1675


DALL-E 2
AI IMAGE
GENERATORS


DALL·E 2
Midjourney v4

DeepMind


$c_{ij} = \sum_{k=1}^{n} a_{ik} \cdot b_{kj}$
Platzi

Ciclo de vida: CRISP-DM moderno

Ciclo de vida: CRISP-DM moderno

- Negocio: problema, valor, métricas
- Datos: EDA, calidad, sesgos
- Preparación: limpieza y *features*
- Modelado: selección y *tuning*
- Evaluación: validación honesta
- Despliegue: MLOps y monitoreo

Reflexión

Sin *baseline* simple, el ML puede no ser necesario.

Validación y particiones

- Train: aprendizaje de parámetros
- Validación: ajuste de hiperparámetros
- Test: estimación imparcial final

$$\text{k-fold CV}: \ \hat{R} = \frac{1}{k}\sum_{i=1}^{k} R_i$$

Pitfall

Fuga de información: no usar test para decidir hiperparámetros.

Métricas esenciales

- Clasificación: Accuracy, Precision, Recall, F1
- Curvas: ROC/AUC, PR para desbalance
- Regresión: MAE, RMSE, R^2

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{Prec} \cdot \mathrm{Rec}}{\mathrm{Prec} + \mathrm{Rec}}$$

Conexión

Métrica técnica debe traducir a KPI de negocio.

Sesgo–varianza y generalización

$$\mathbb{E}[(y-\hat{f}(x))^2] = \underbrace{\text{Bias}^2}_{\text{subajuste}} + \underbrace{\text{Var}}_{\text{sobreajuste}} + \sigma^2$$

- Más complejidad $\downarrow$ sesgo, $\uparrow$ varianza
- Regularizar, simplificar o recolectar más datos

Teorema clave

Compromiso sesgo–varianza determina el error fuera de muestra.

No Free Lunch (NFL)

- Ningún algoritmo domina en todos los problemas
- Rendimiento depende de supuestos y datos
- Modelado: comparar y validar empíricamente

$$\forall \mathcal{A},\ \exists\ \text{tarea}:\ \ \mathcal{A}\ \text{no es óptimo}$$

Implicación

Evitar recetas universales; diseñar para el contexto.

Drift y monitoreo en producción

- *Covariate shift*: cambia $p(X)$
- *Concept drift*: cambia $p(Y|X)$
- *Prior shift*: cambia $p(Y)$

$$d(p_{\text{prod}}, p_{\text{train}}) > \tau \;\Rightarrow\; \text{reentrenar}$$

Buenas prácticas

Alertas, A/B testing, *canary*, *shadow* deployment.

Errores comunes y buenas prácticas

- Métrica inadecuada para el objetivo
- *Leakage* entre train/valid/test
- Sin *baseline* y ablatión
- Ignorar desbalance y costo de errores
- Falta de documentación y trazabilidad

Checklist mínimo

Datos limpios, CV estratificada, *pipeline* reproducible.

Síntesis: lo esencial

- Problema de negocio guía todo el diseño
- Datos superan a complejidad del modelo
- Validación rigurosa evita autoengaño
- Sesgo–varianza: equilibrio o fracaso
- Producción requiere monitoreo continuo

Referencias y lecturas recomendadas I

- Mitchell, *Machine Learning* (1997)
- Bishop, *PRML* (2006)
- Hastie, Tibshirani, Friedman, *ESL* (2009)
- Goodfellow, Bengio, Courville, *Deep Learning* (2016)
- LeCun, Bengio, Hinton, *Deep Learning*, Nature (2015)
- Burkov, *Machine Learning Engineering* (2020)
- O'Neil, *Weapons of Math Destruction* (2016)
- Christian, *The Alignment Problem* (2020)
- Distill.pub, Papers with Code, arXiv cs.LG

¡Muchas gracias por su atención!

¿Preguntas?

Contacto: Marco Teran
webpage: marcoteran.github.io/
e-mail: mtteranl@eafit.edu.co