

Introducción a Machine Learning

Aprendizaje Automatico

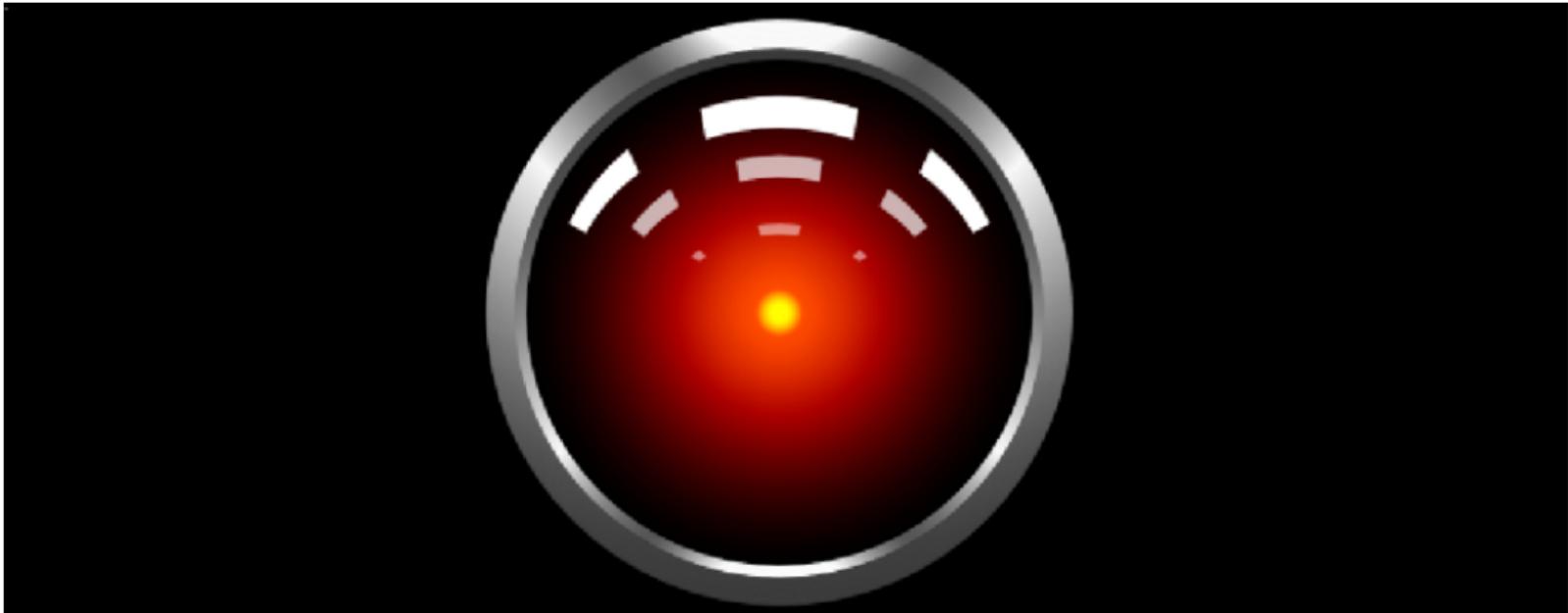


Marco Teran
EAFIT

2025

Contenido

- 1 Objetivos y mapa
- 2 Un poco de historia...
- 3 ¿Qué es la Inteligencia Artificial?
- 4 Aprendizaje Computacional
 - Redes Neuronales
- 5 Aplicaciones
- 6 Ciclo de vida: CRISP-DM moderno



► 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 Gáspárav (1997)

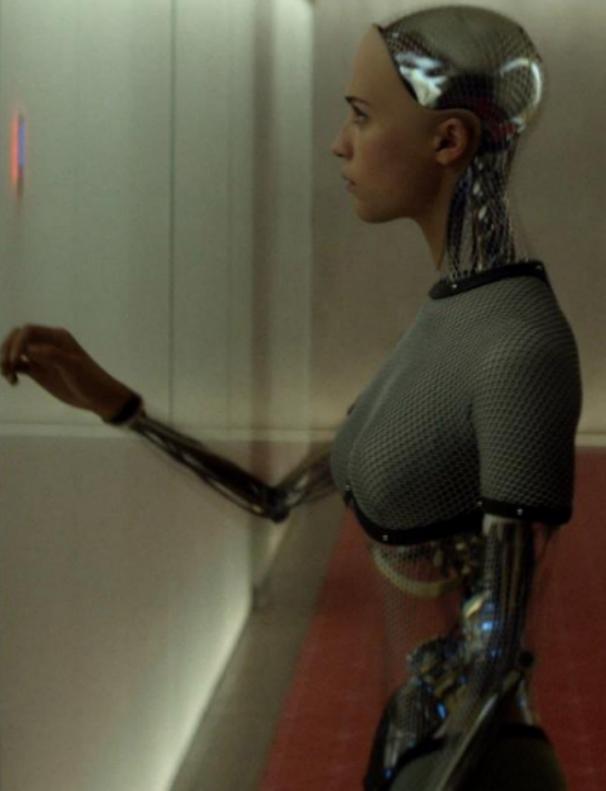


► ver video

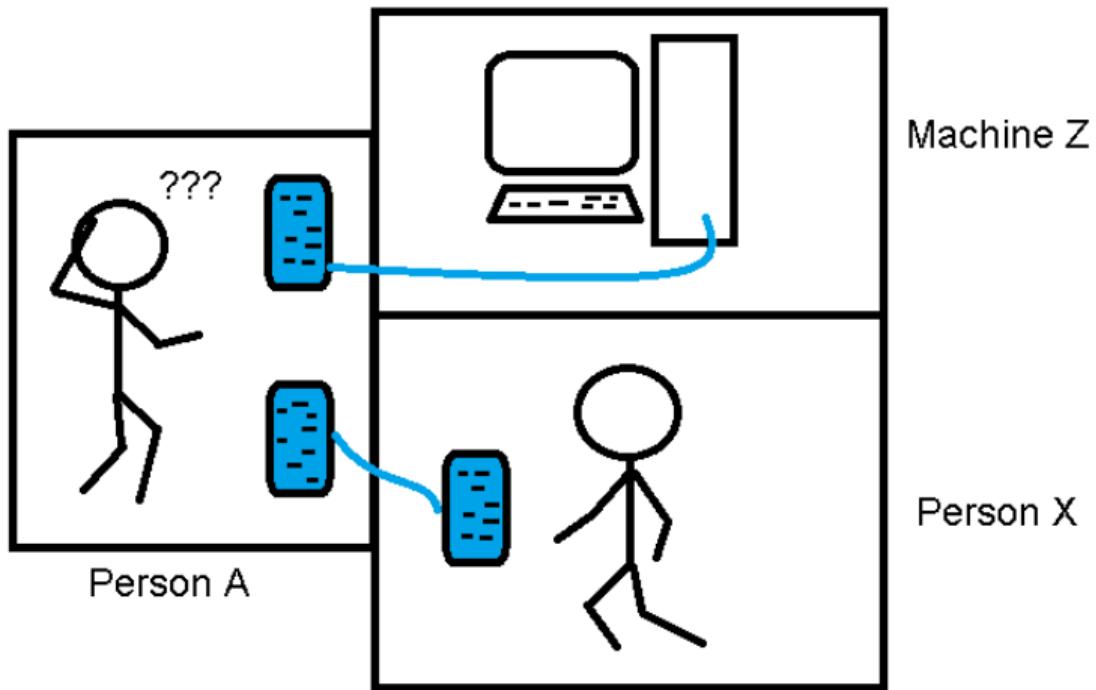
computer chess



¿Qué es la Inteligencia Artificial?

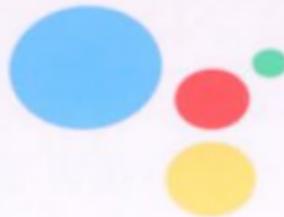


El test de Turing



El test de Turing





Hi, how can I help?



Google Duplex (2018)



Google Duplex

Advancing AI for Everyone



► ver video

Inteligencia y racionalidad computacional

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

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

Conexión

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

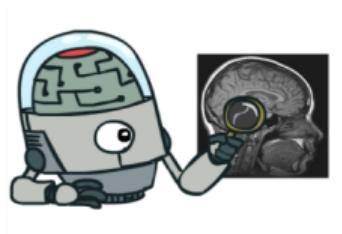
Inteligencia Artificial

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

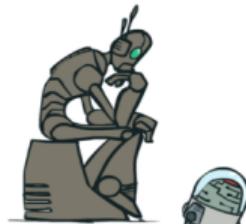
“the ability to take the right decisions, according to some criterion
(e.g. survival and reproduction, for most animals)”

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

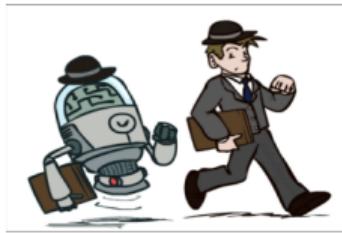
Cuatro enfoques clásicos de IA



(a) Pensar como humano



(b) Pensar razonadamente



(c) Actuar como humano



(d) Actuar razonadamente

¿Qué es IA?

La ciencia de crear máquinas que:

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

Agente: software inteligente.

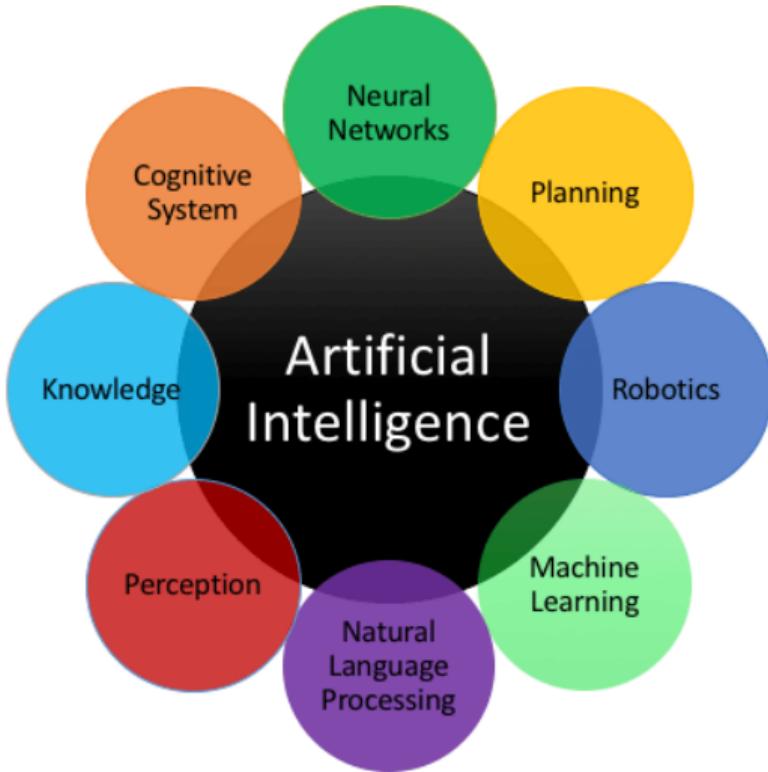
Robot: plataforma física de acción.

IA hoy: racionalidad computacional

- Objetivos $\$ \rightarrow \$$ función de utilidad
- Agir racional $\$ \rightarrow \$$ 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

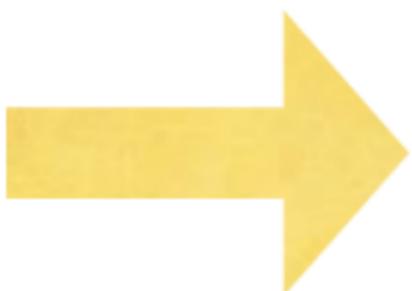
1997

1994

1968

1994

1995



1993

1997

1994

1968

1994

1945

Aprendizaje de máquina vs programación tradicional

Programación tradicional



Machine Learning



Reglas (código) \Rightarrow salidas vs Datos \Rightarrow modelo f

Aprendizaje de máquina: definición operativa

“The acquisition of knowledge or skills through study or experience”

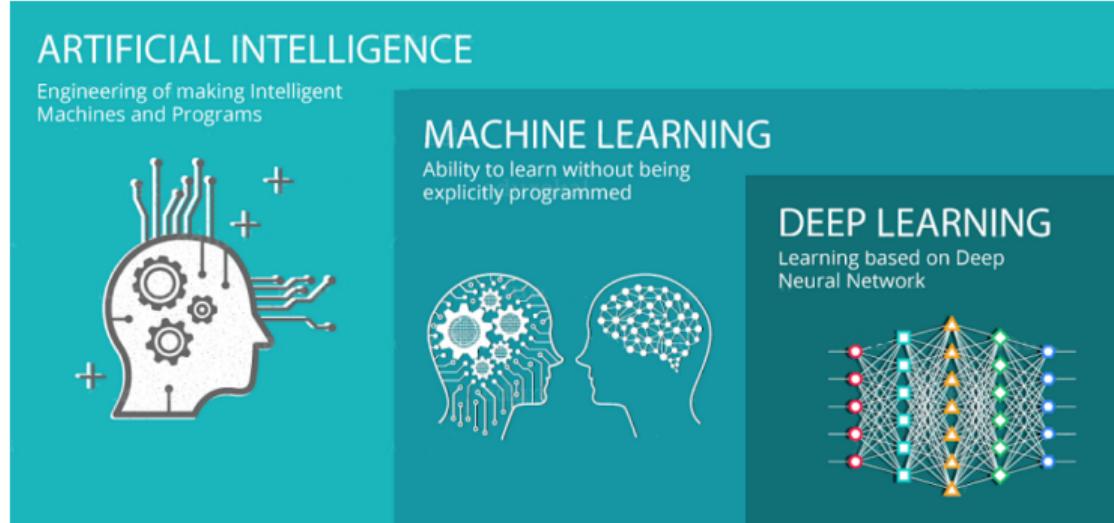
El **machine learning** estudia algoritmos que permiten a un agente *aprender a decidir/actuar* desde datos.

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

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

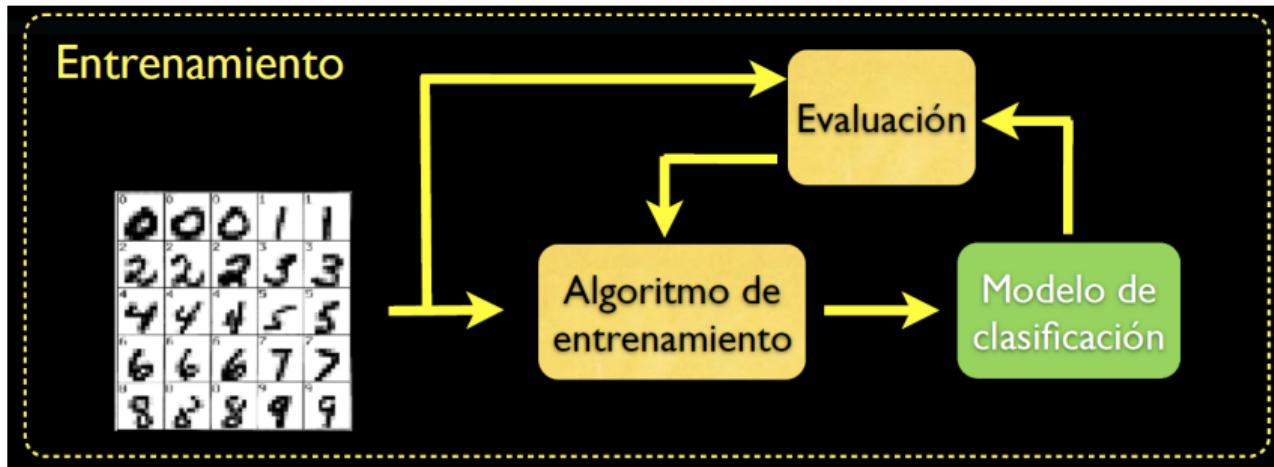
A program that can sense, reason,
act, and adapt

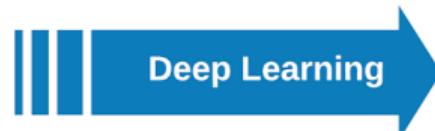
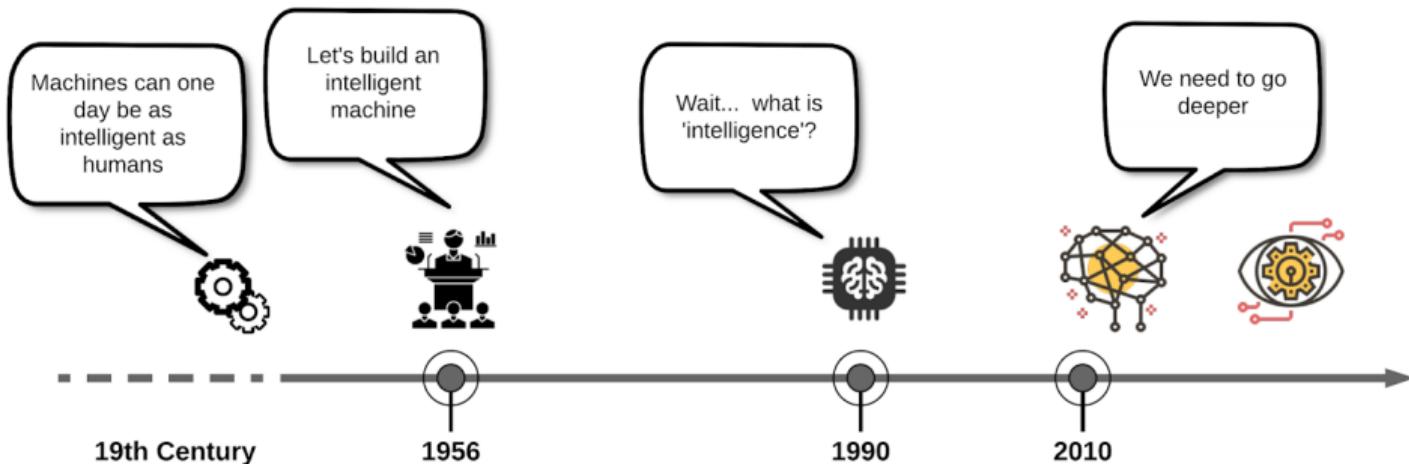
MACHINE LEARNING

Algorithms whose performance improve
as they are exposed to more data over time

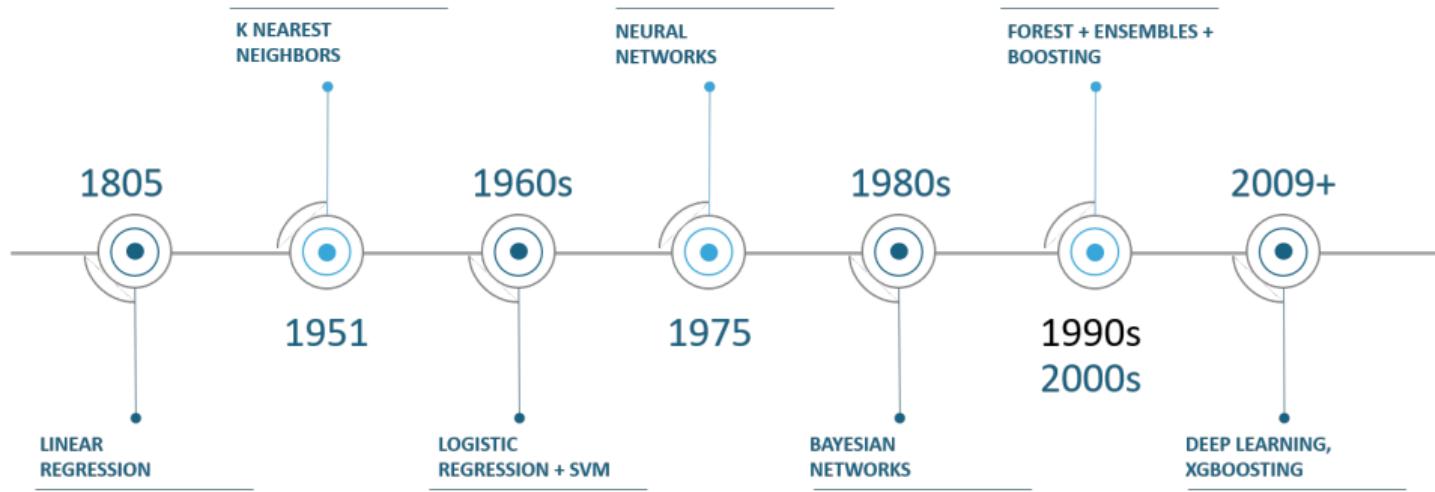
DEEP LEARNING

Subset of machine learning in
which multilayered neural
networks learn from
vast amounts of data





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

Supervisado : (x, y) 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

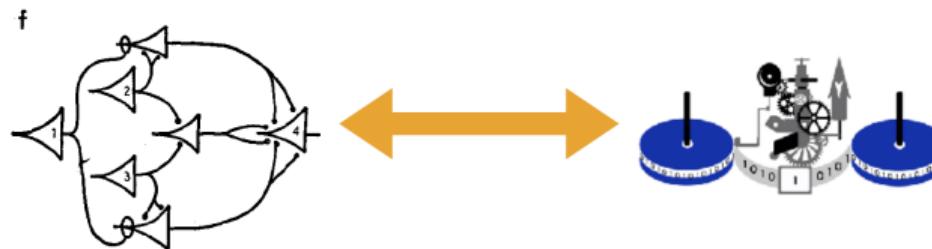
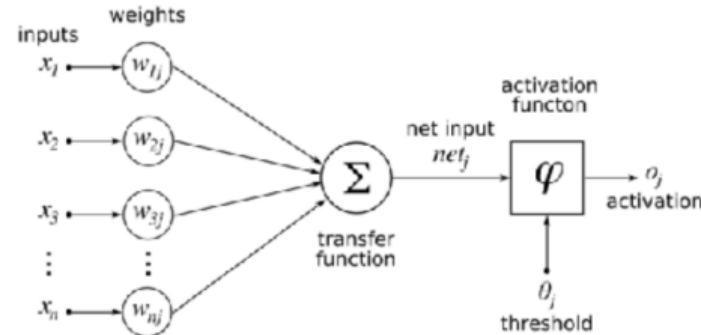
McCulloch & Pitts Artificial Neuron

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

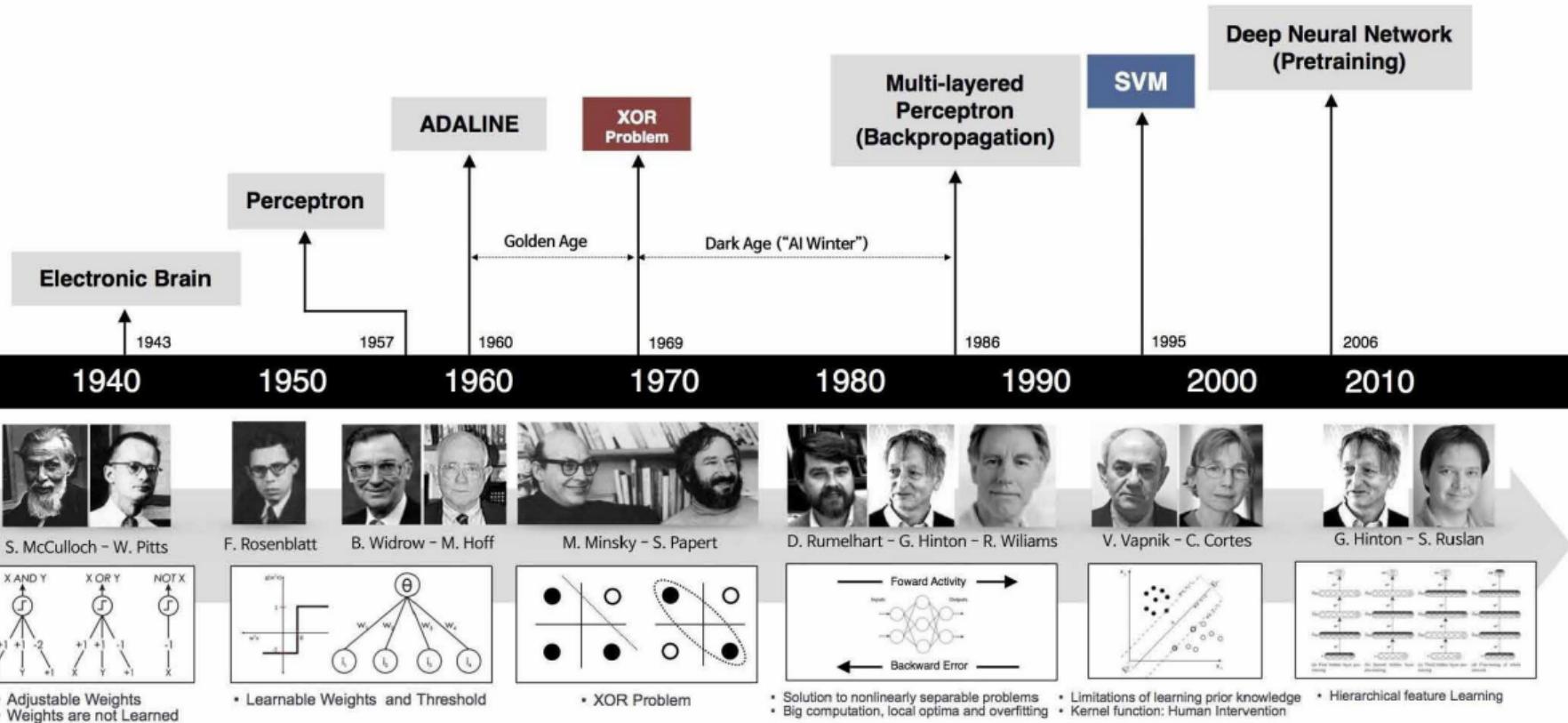


Machine Learning

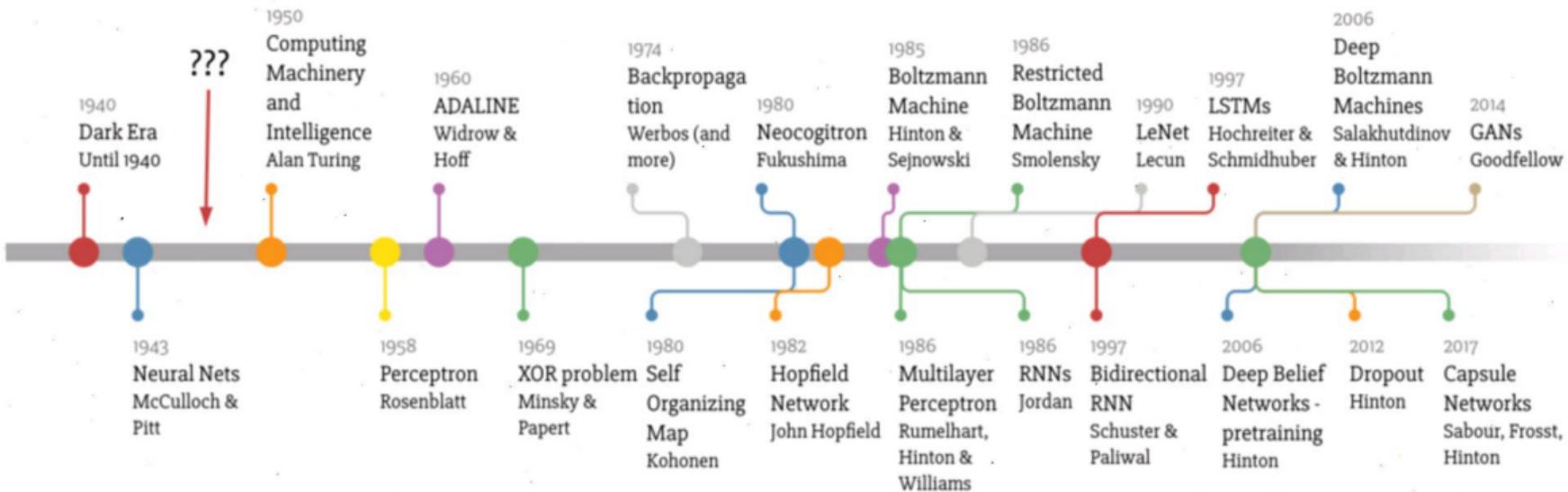


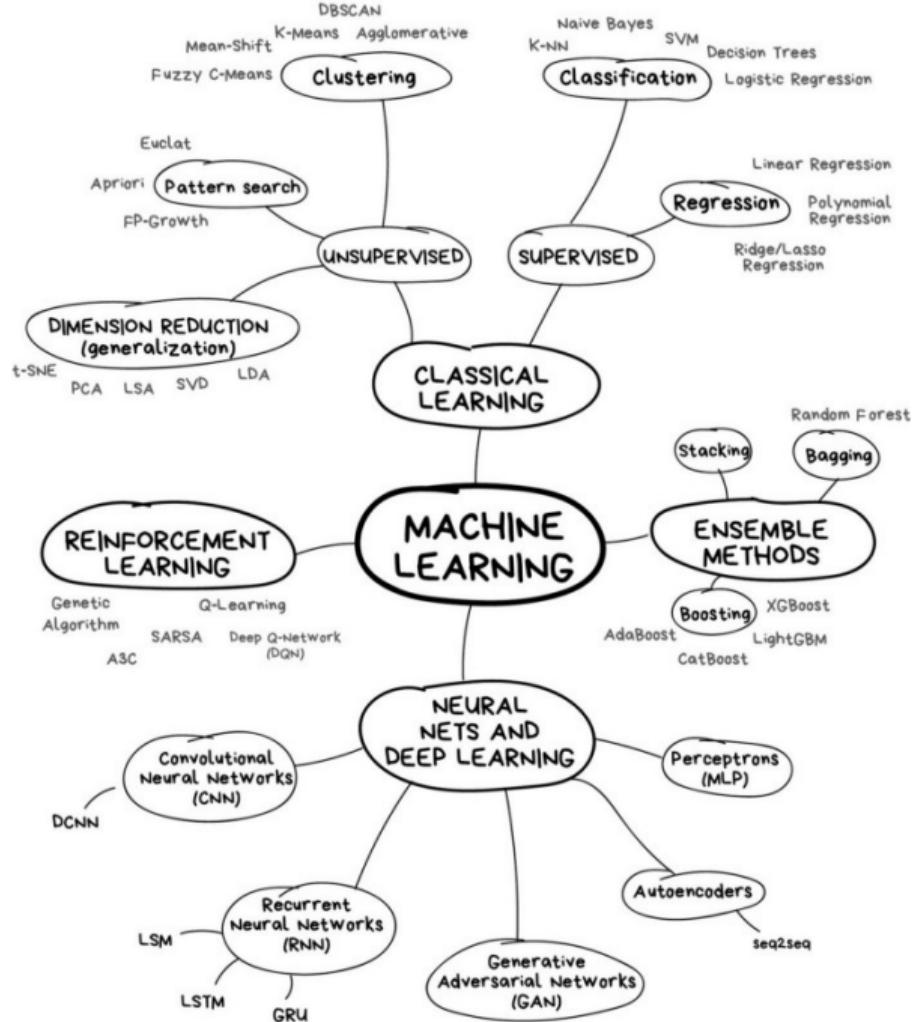
Deep Learning

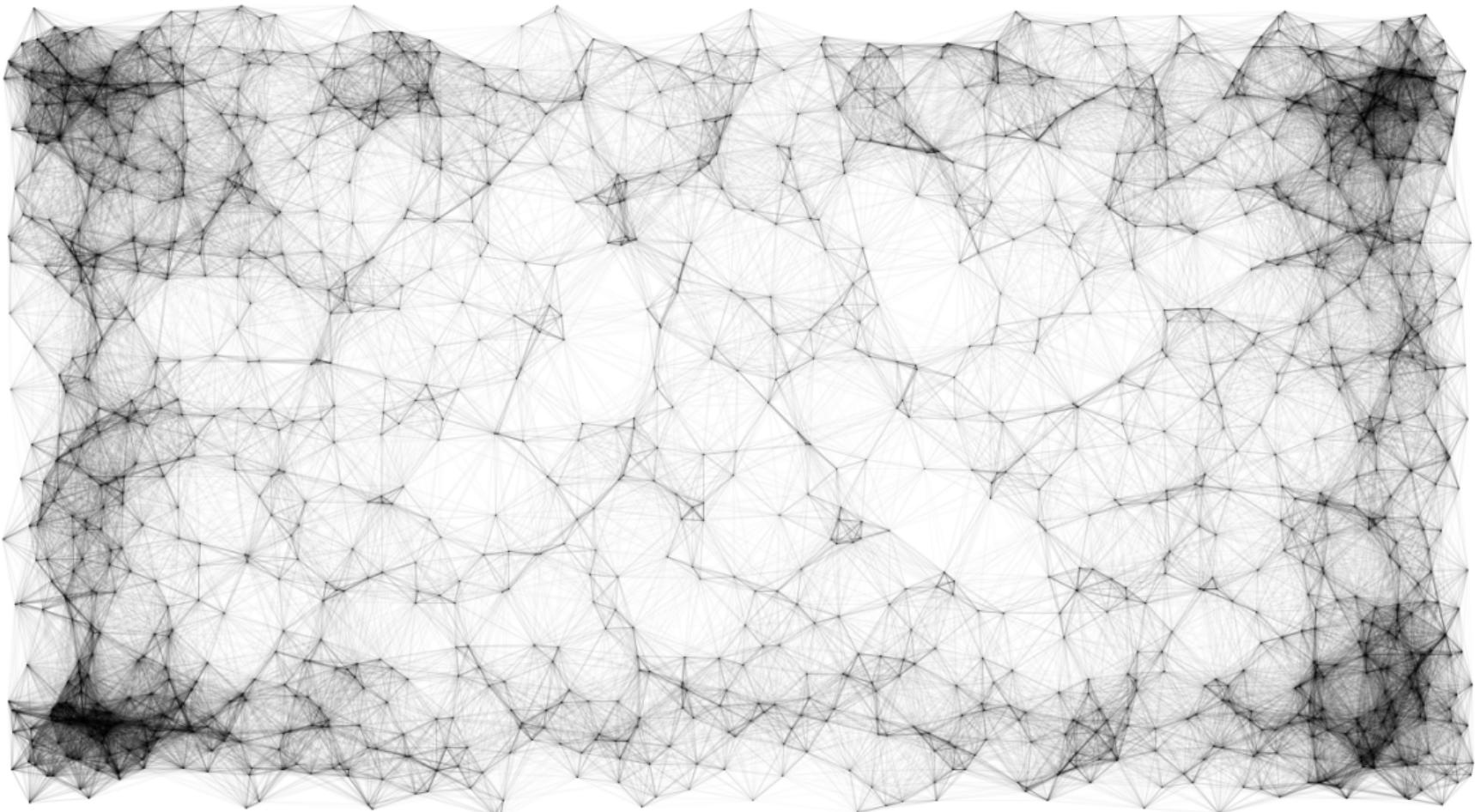




Deep Learning Timeline



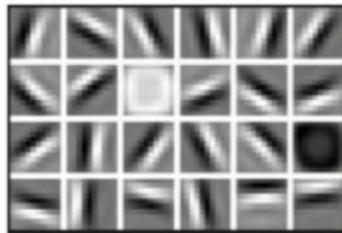




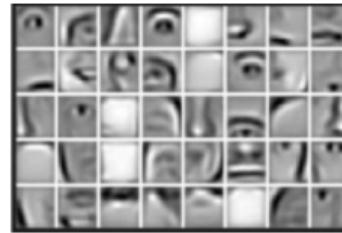
¿Por qué Deep Learning y por qué ahora?

¿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

1952	Stochastic Gradient Descent
1958	Perceptron <ul style="list-style-type: none">• Learnable Weights
⋮	⋮
1986	Backpropagation <ul style="list-style-type: none">• Multi-Layer Perceptron
1995	Deep Convolutional NN <ul style="list-style-type: none">• Digit Recognition
⋮	⋮

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

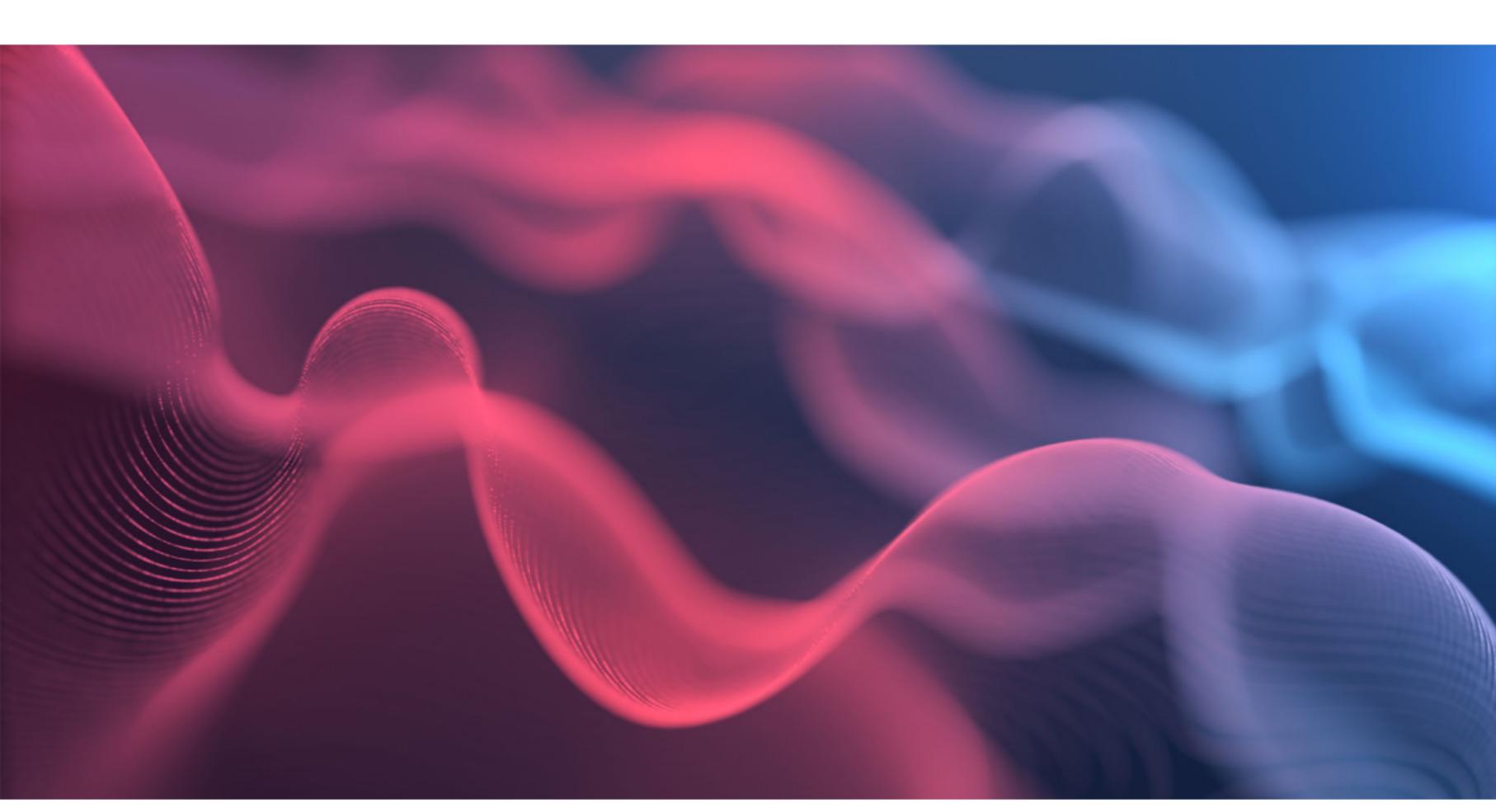
- Unidades de procesamiento gráfico (GPU)
- Masivamente paralelizable



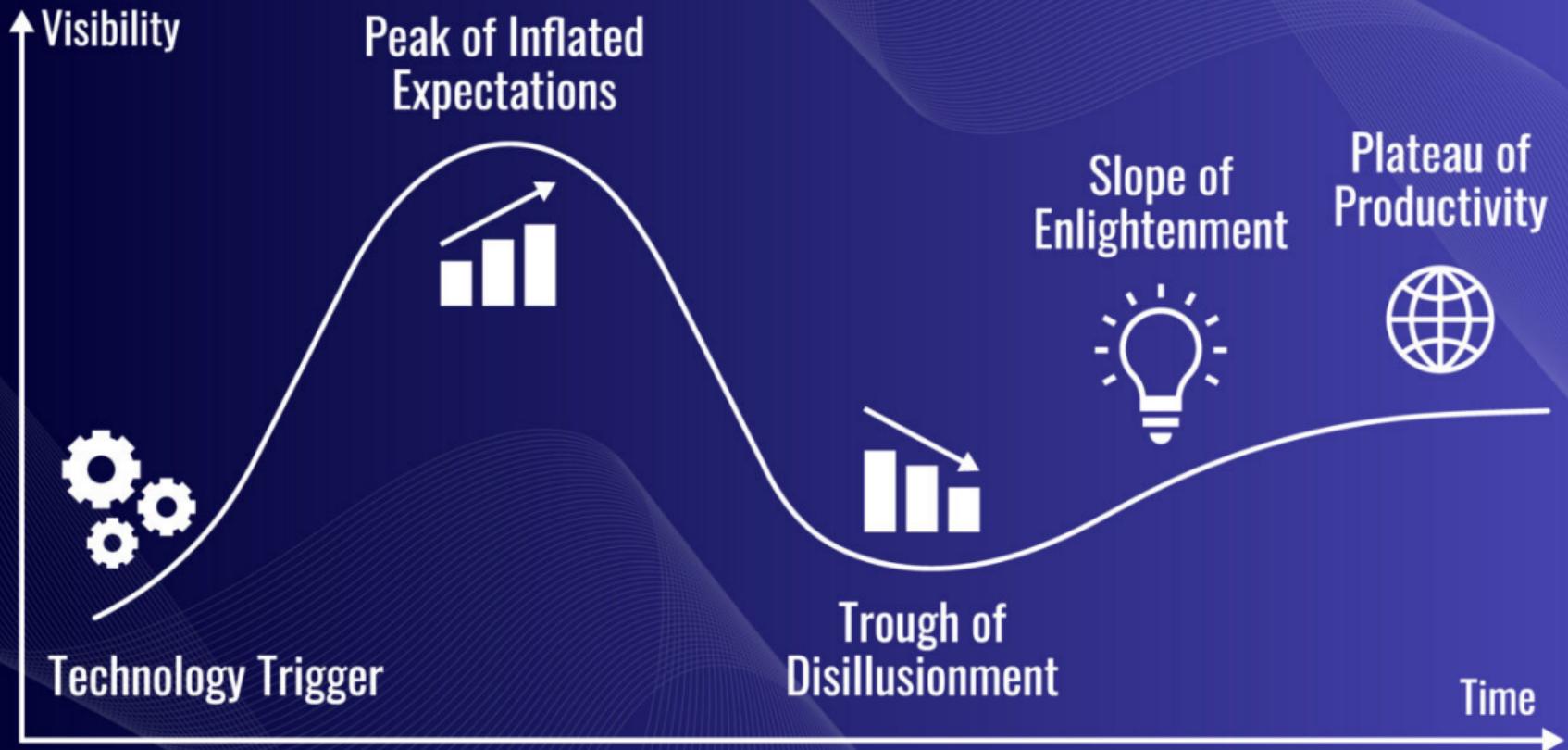
3. Software

- Técnicas mejoradas
- Nuevos modelos
- Toolboxes

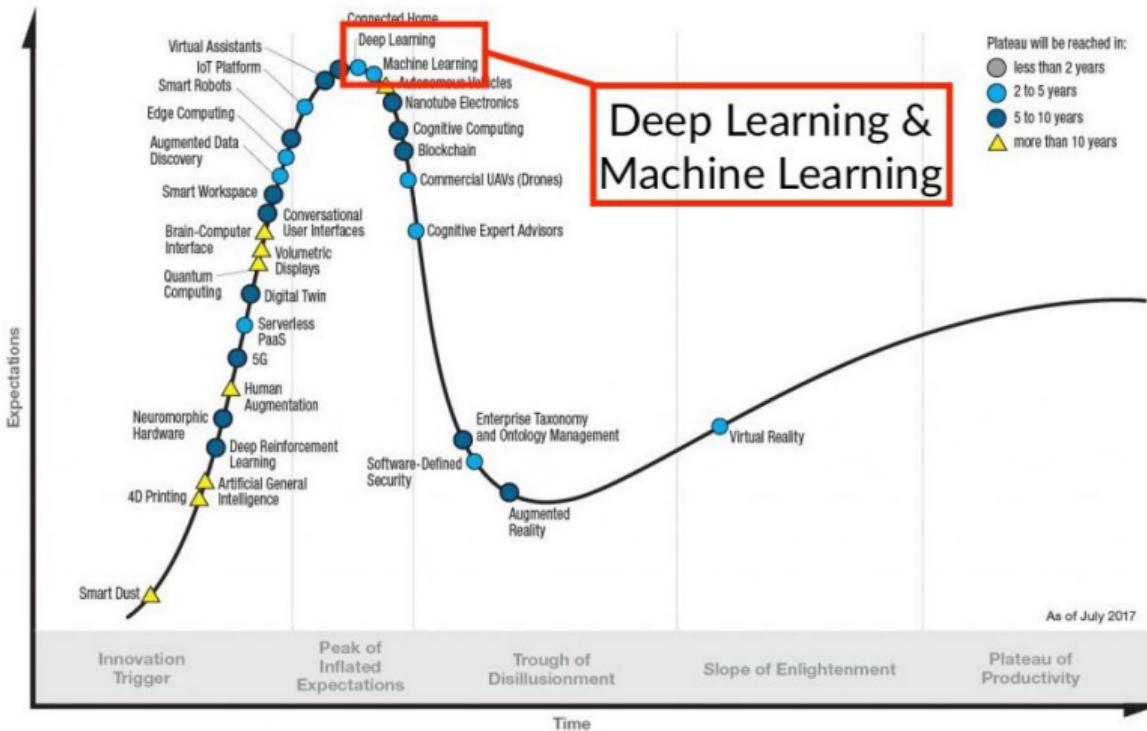




Aplicaciones



Gartner Hype Cycle for Emerging Technologies, 2017



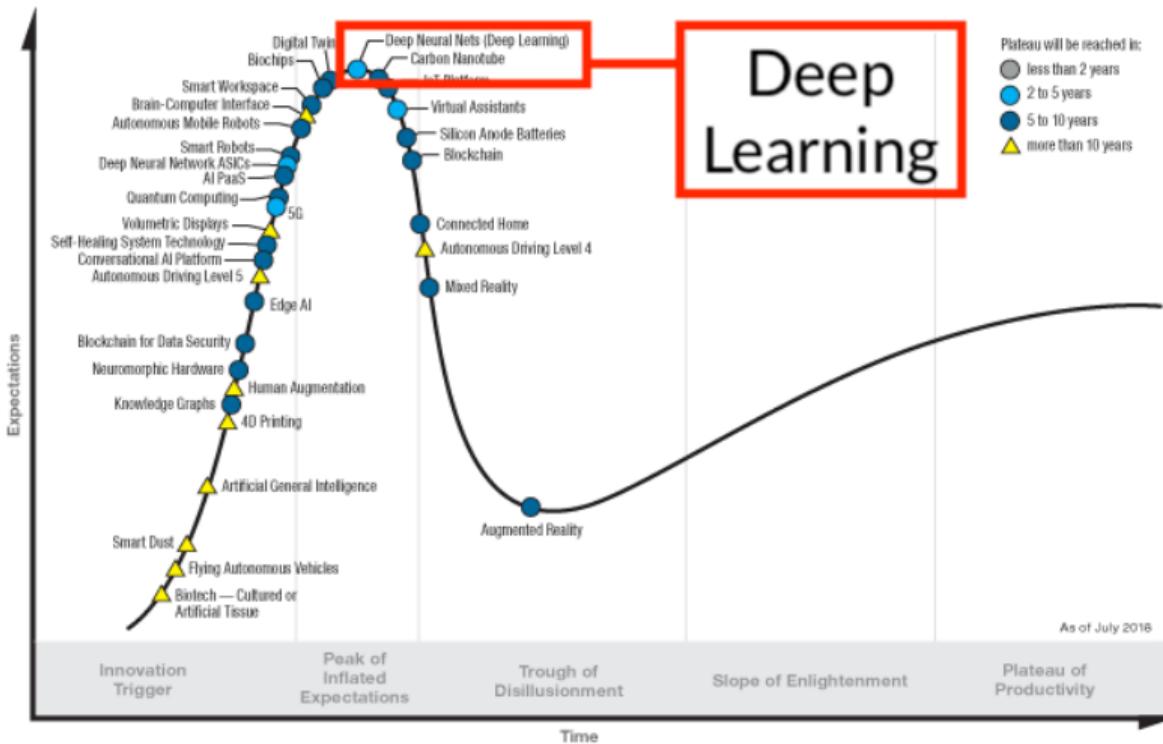
gartner.com/SmarterWithGartner

Source: Gartner (July 2017)

© 2017 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

Hype Cycle for Emerging Technologies, 2018

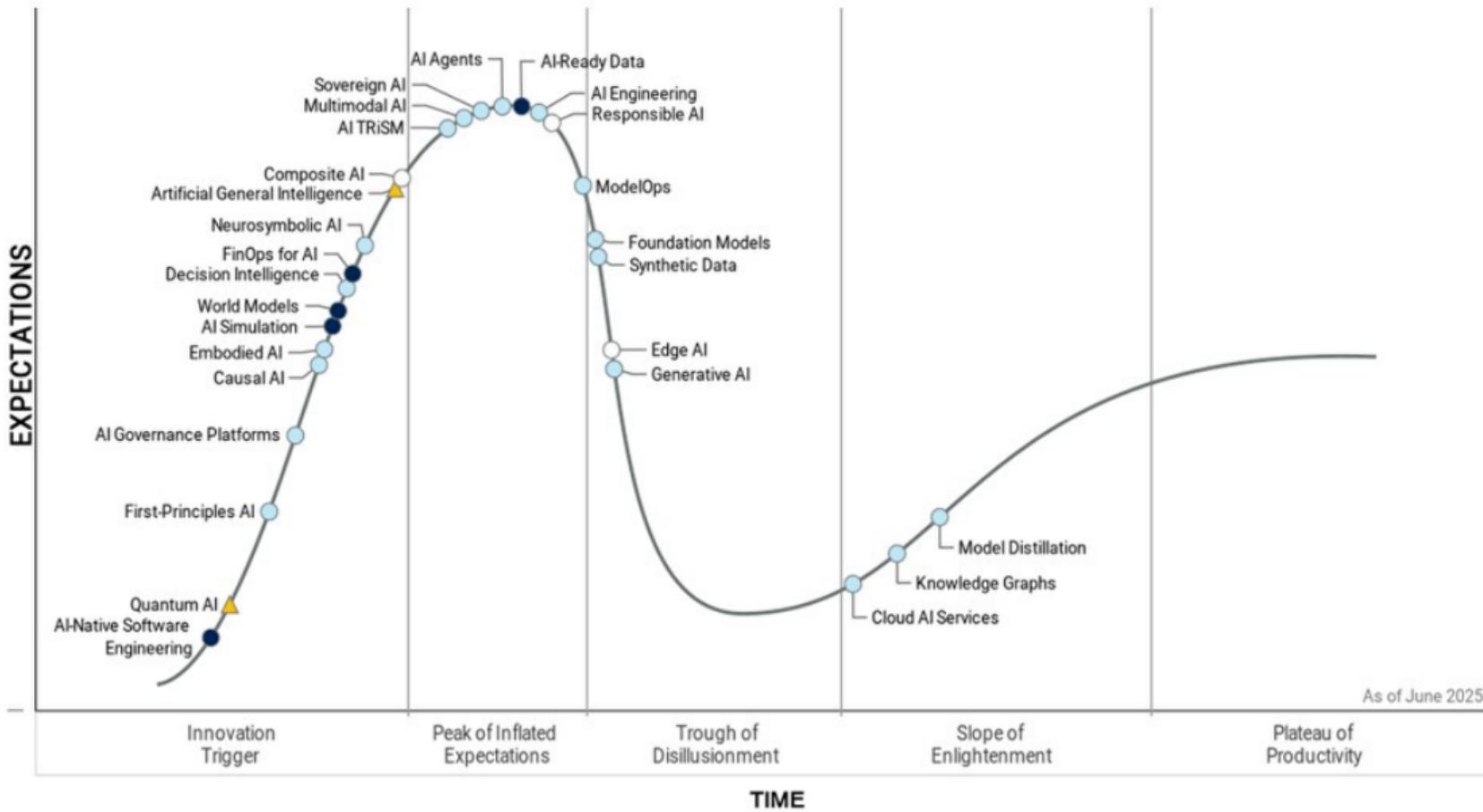


gartner.com/SmarterWithGartner

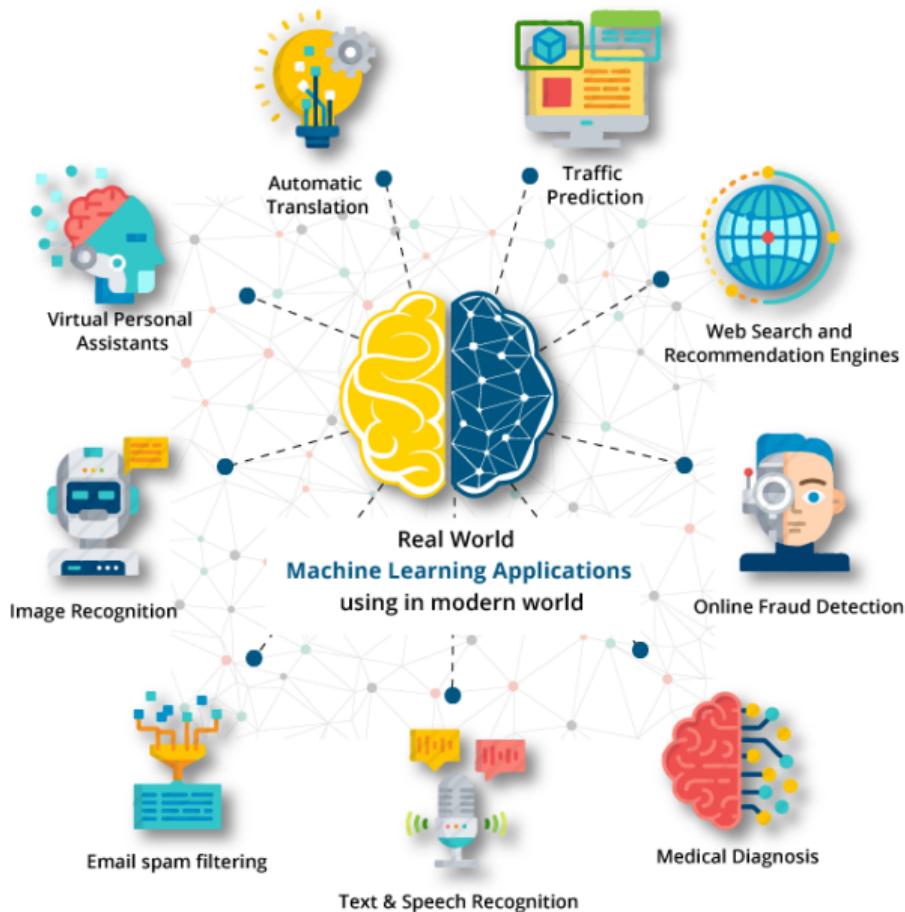
Source: Gartner (August 2018)

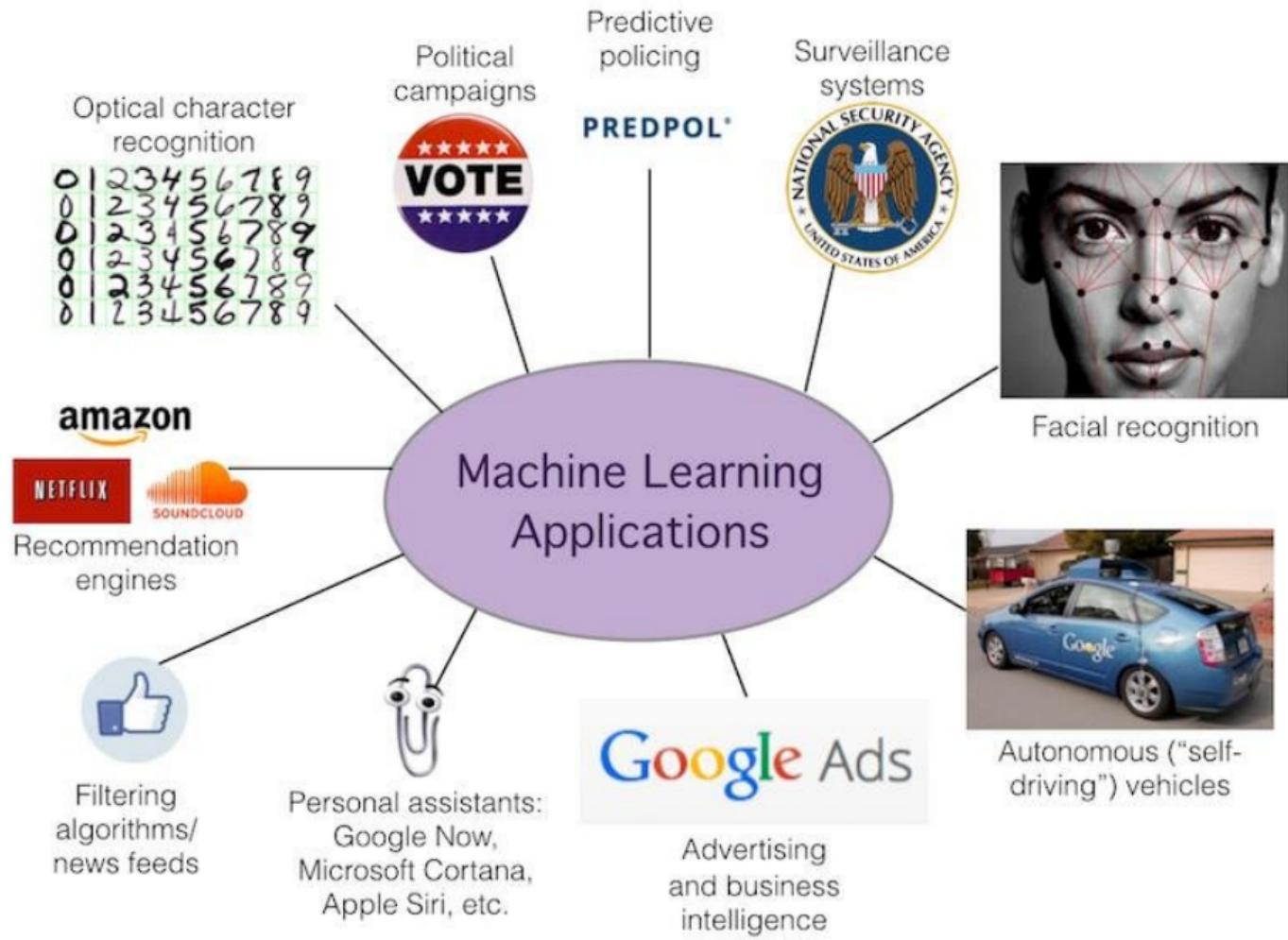
© 2018 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner



Plateau will be reached: ○ < 2 yrs. ● 2–5 yrs. ● 5–10 yrs. ▲ >10 yrs. ✗ Obsolete before plateau





Procesamiento de lenguaje natural

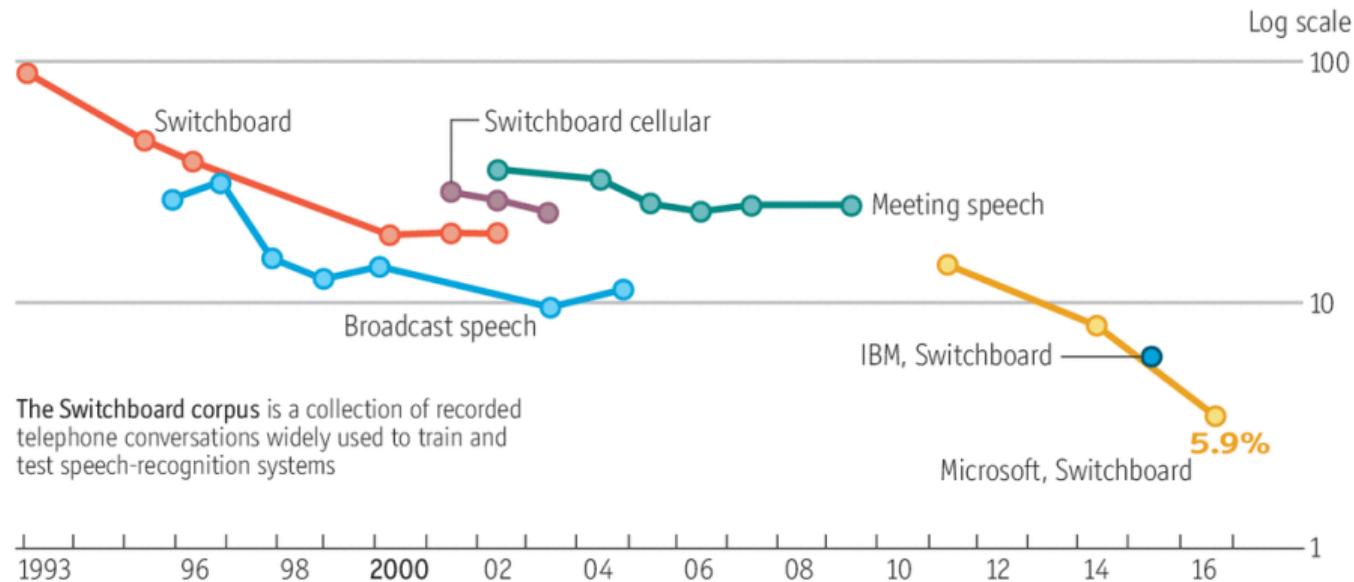
- Dado un texto, predecir la temática
- Dado un email, predecir si este es un spam
- Dado un texto, predecir el idioma y su traducción a otro lenguaje



Reconocimiento de Voz

Loud and clear

Speech-recognition word-error rate, selected benchmarks, %



The **Switchboard corpus** is a collection of recorded telephone conversations widely used to train and test speech-recognition systems

Sources: Microsoft; research papers

Economist.com

Reconocimiento de Voz

FAR-FIELD

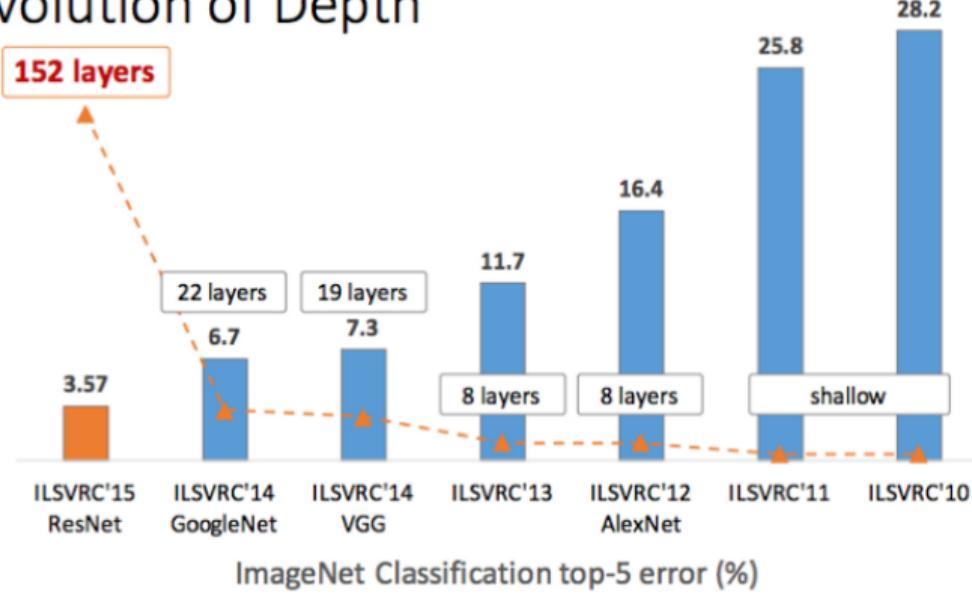
VOICE RECOGNITION



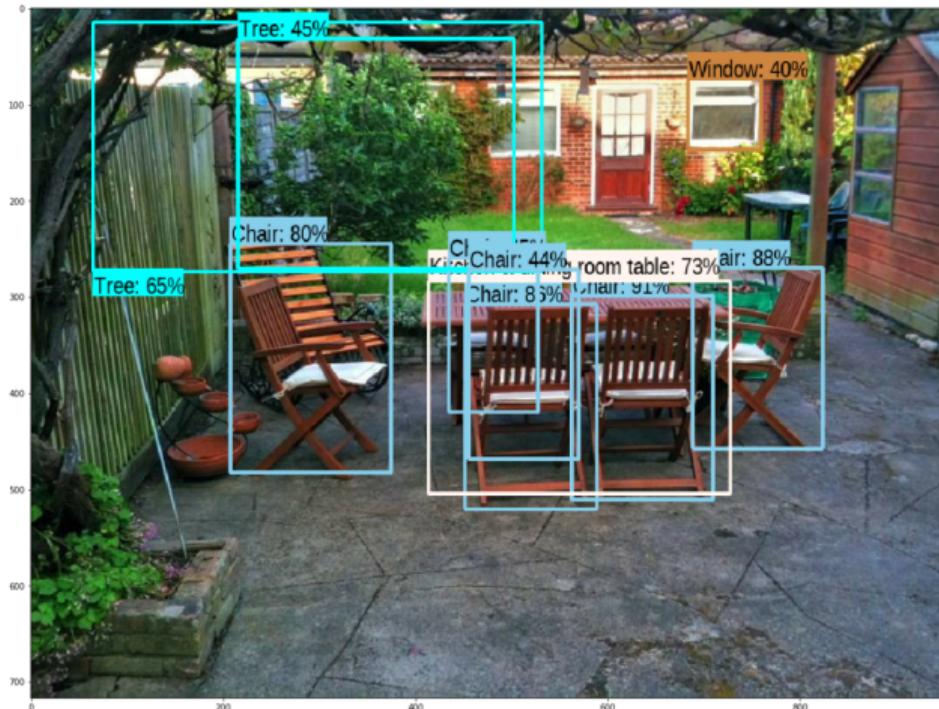
► ver video

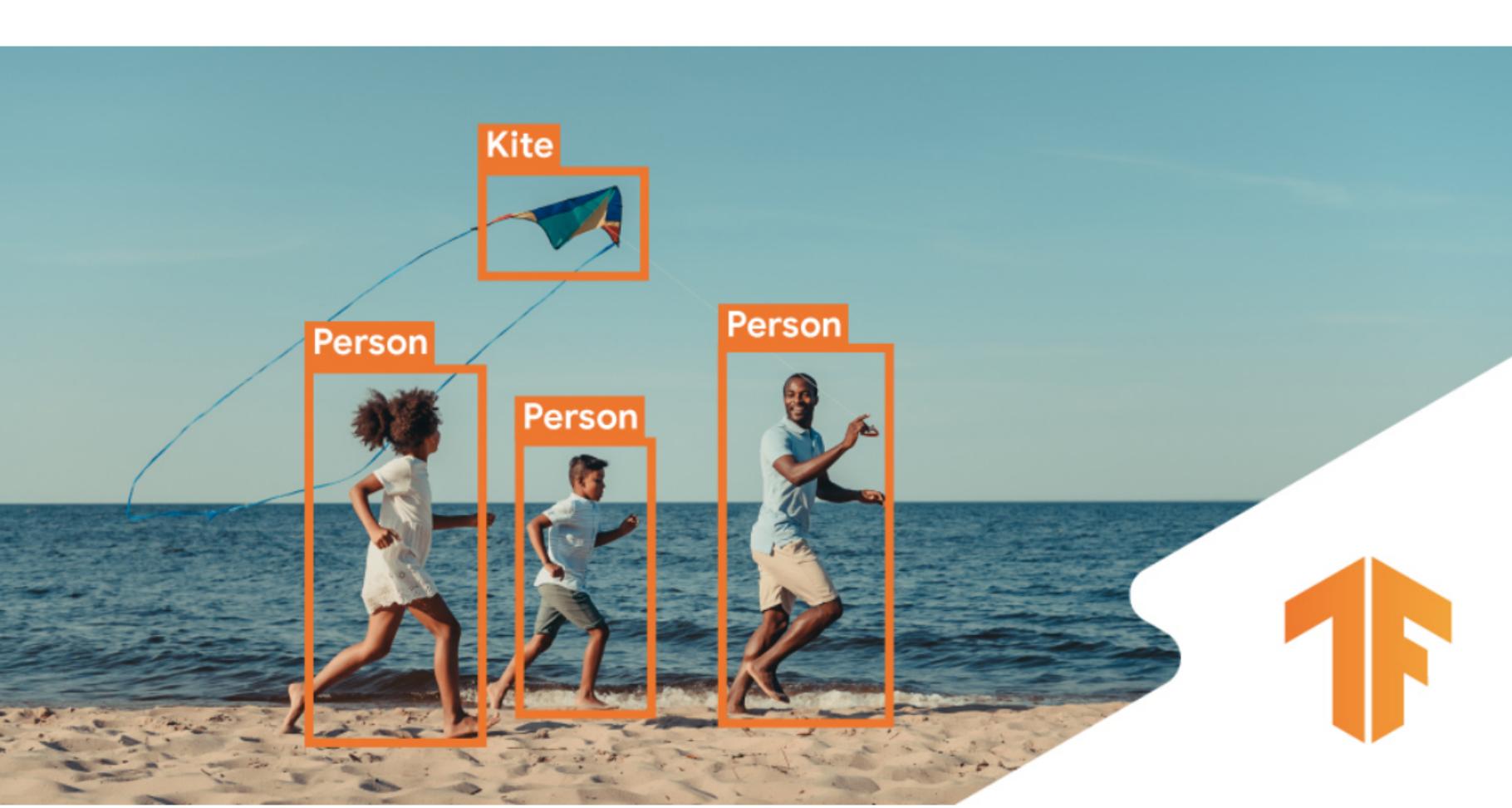
Visión por Computador

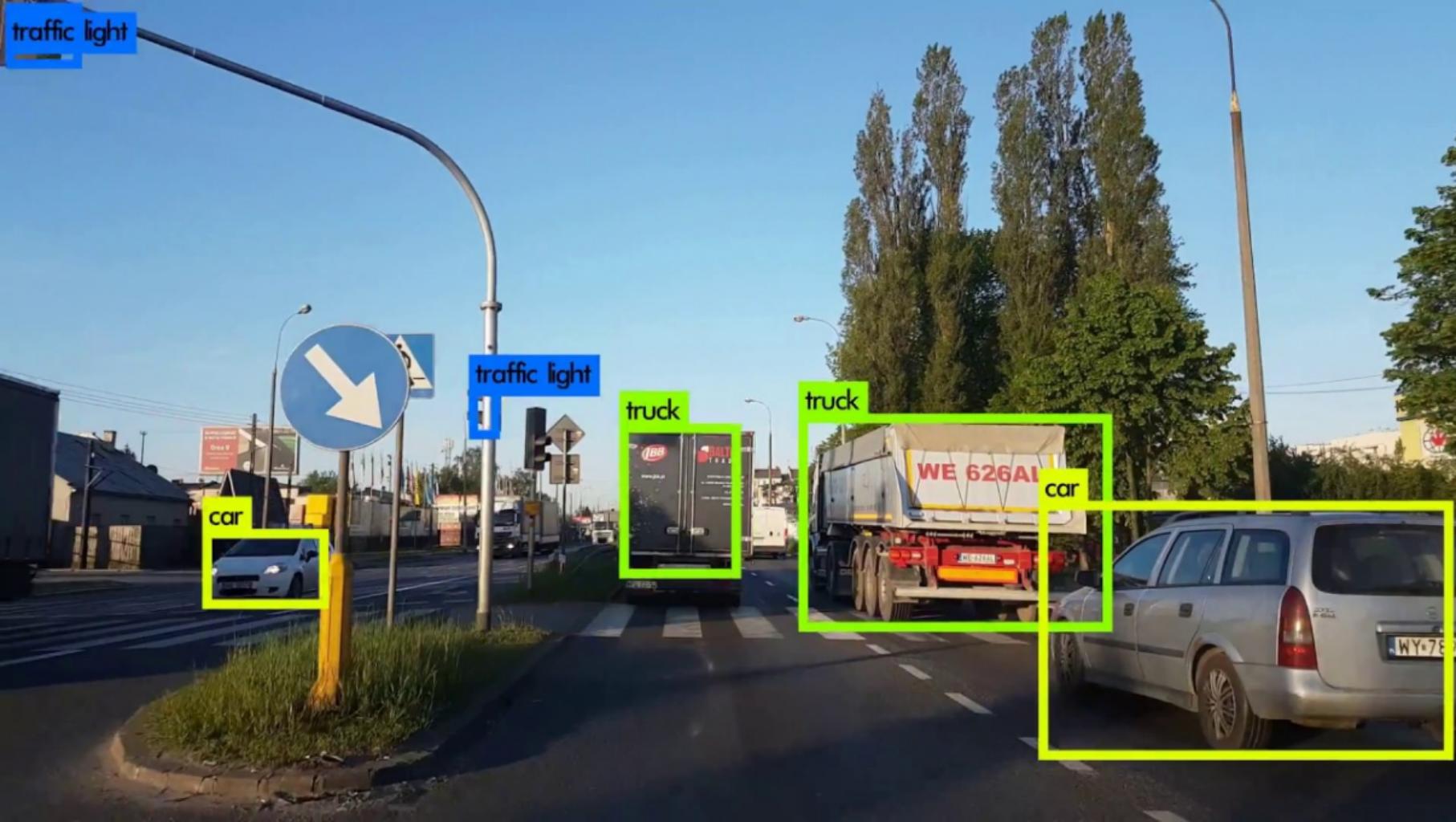
Revolution of Depth



Detección de objetos





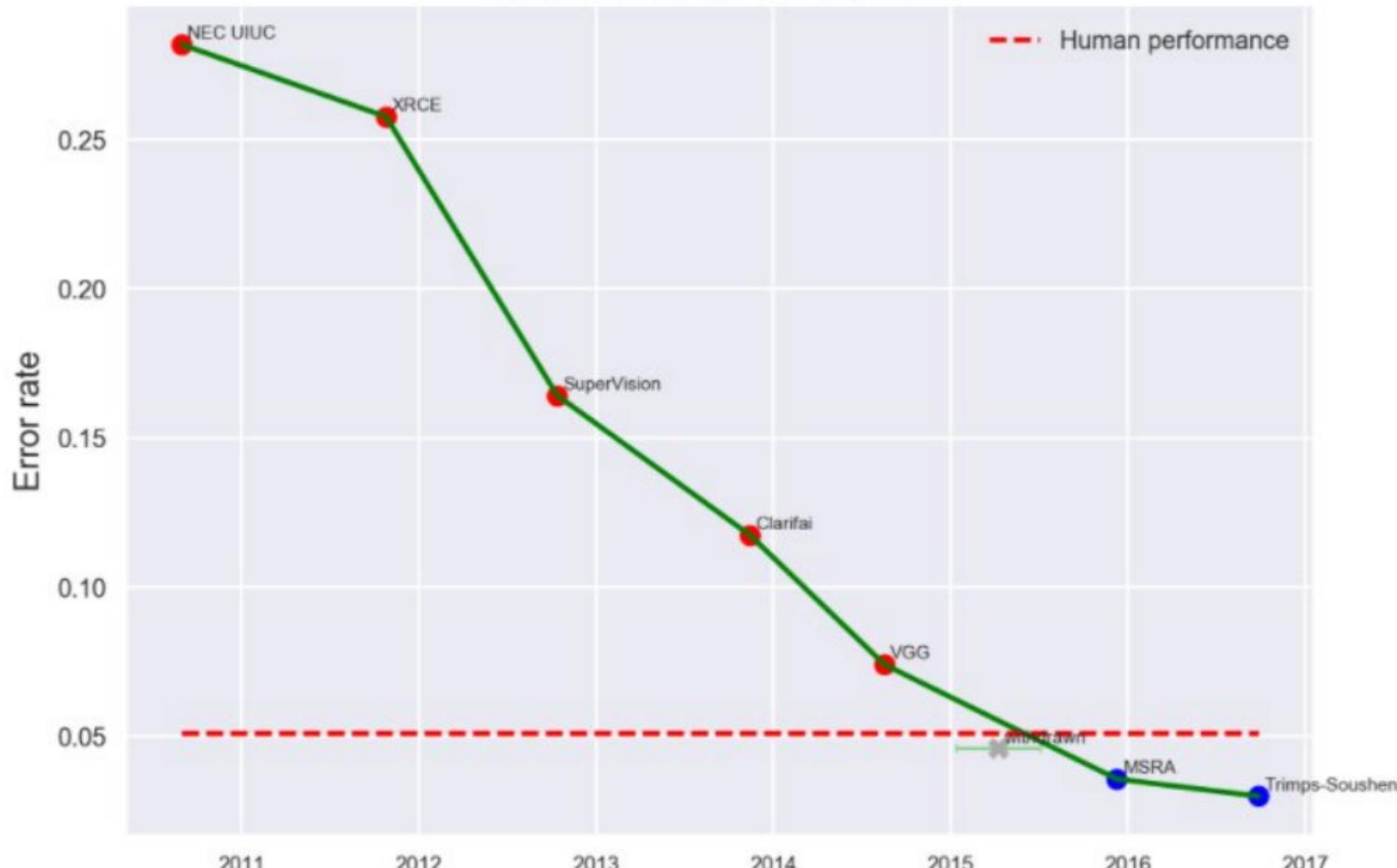


Detección de objetos



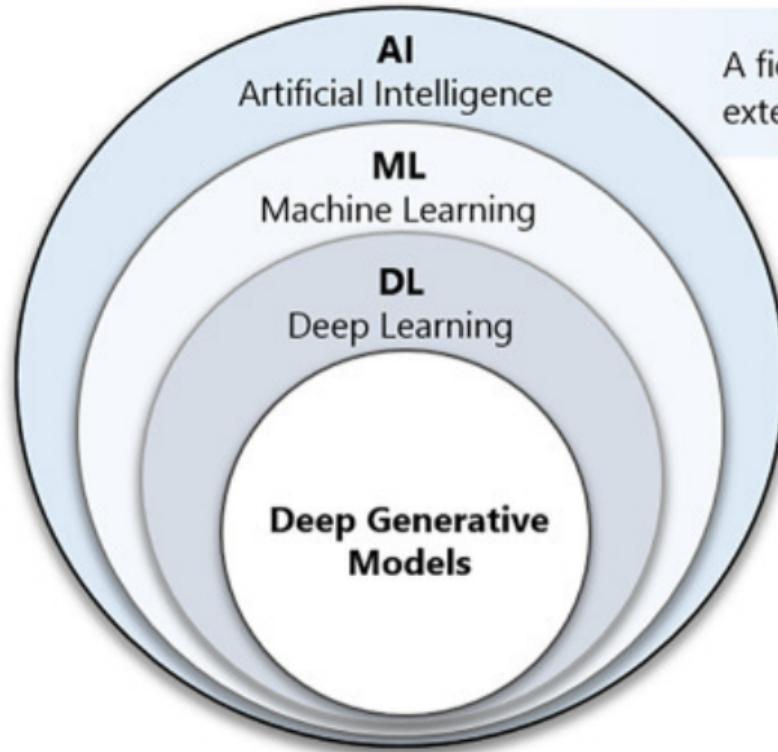
▶ demo

Imagenet Image Recognition



Semantic Segmentation





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

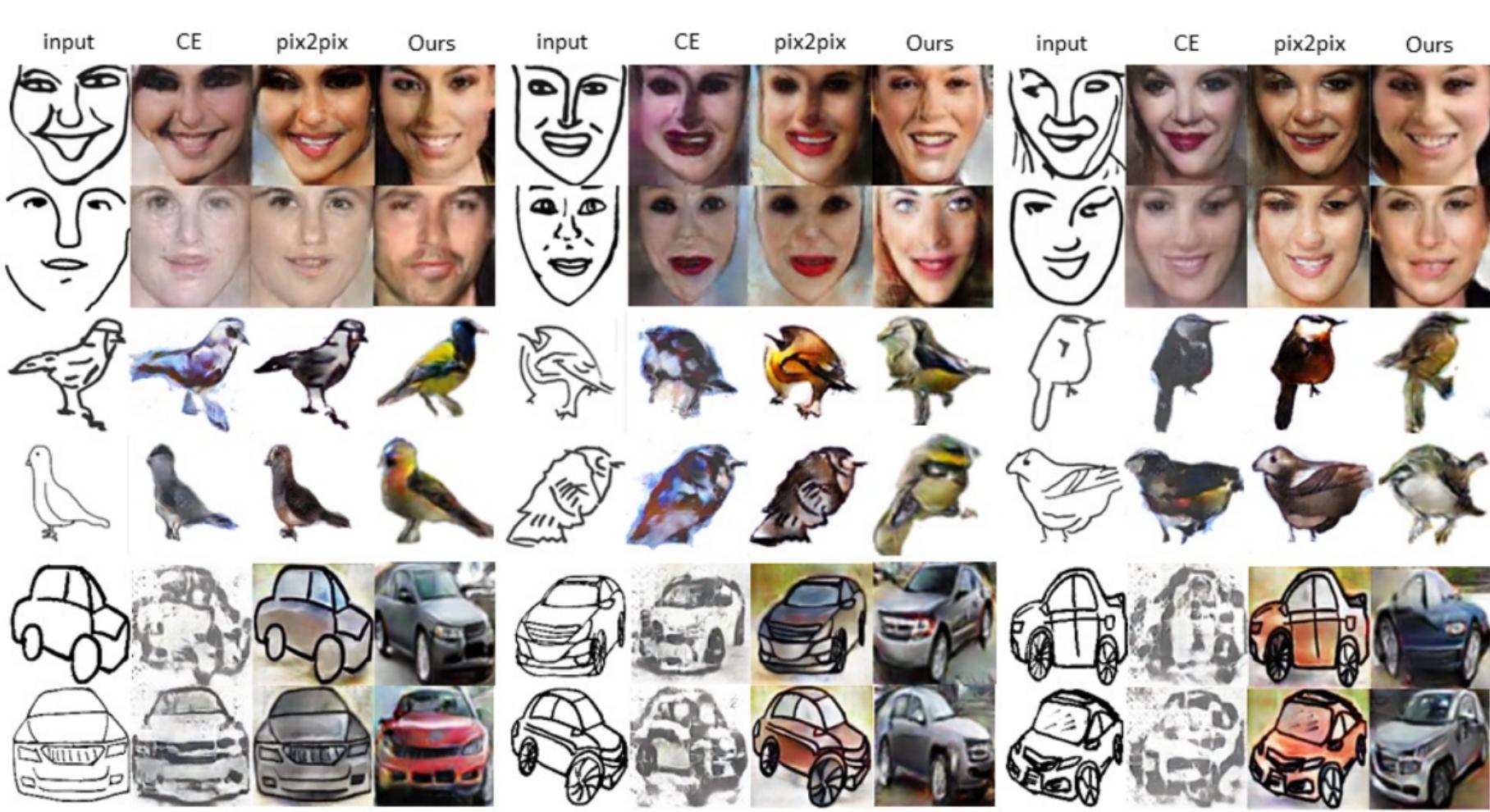
- | | |
|------------------|--|
| Explicit Density | <ul style="list-style-type: none">• Pixel RNN• Variational Autoencoder |
| Implicit Density | <ul style="list-style-type: none">• Generative Adversarial Network• GSN |

Style transfer

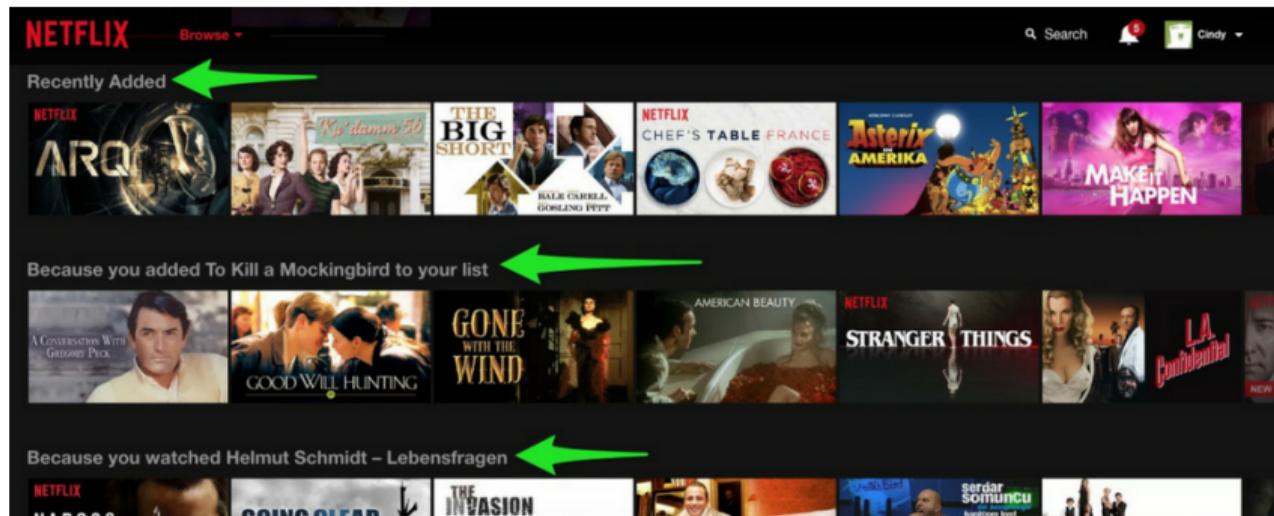
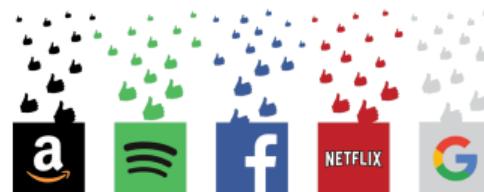


Style transfer





Recommender systems



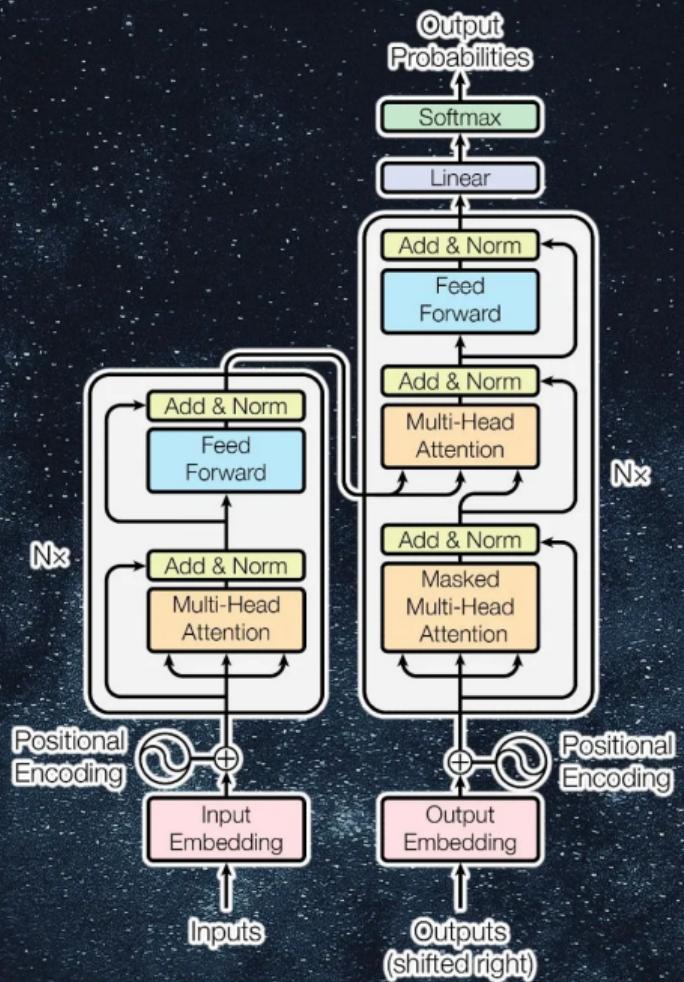
The screenshot shows the Netflix homepage with three examples of recommendation systems:

- Recently Added:** An arrow points to the first row of movie and TV show thumbnails, labeled "Recently Added".
- Because you added To Kill a Mockingbird to your list:** An arrow points to the second row, labeled "Because you added To Kill a Mockingbird to your list". This is a personalized recommendation.
- Because you watched Helmut Schmidt – Lebensfragen:** An arrow points to the third row, labeled "Because you watched Helmut Schmidt – Lebensfragen". This is a content-based recommendation.

Coarse styles copied



Transformers: La nueva joya del aprendizaje profundo



GPT-3

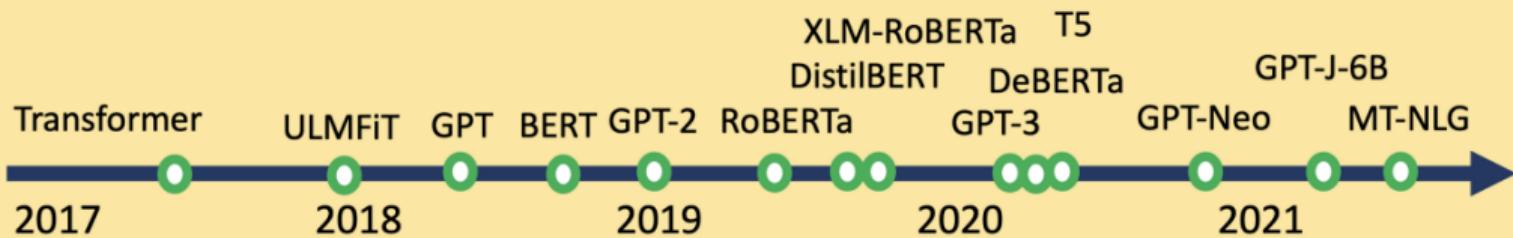
DALL·E

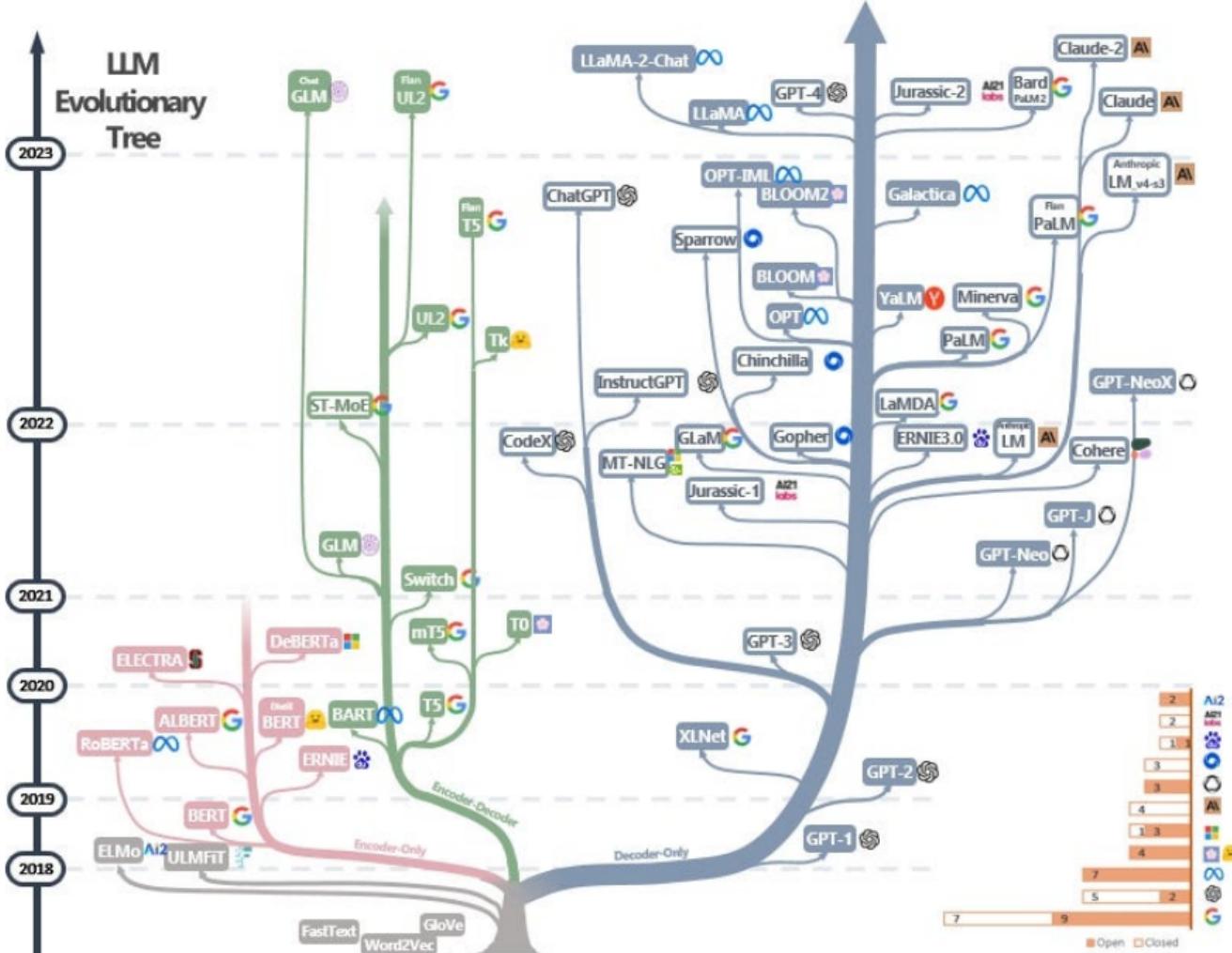
BERT

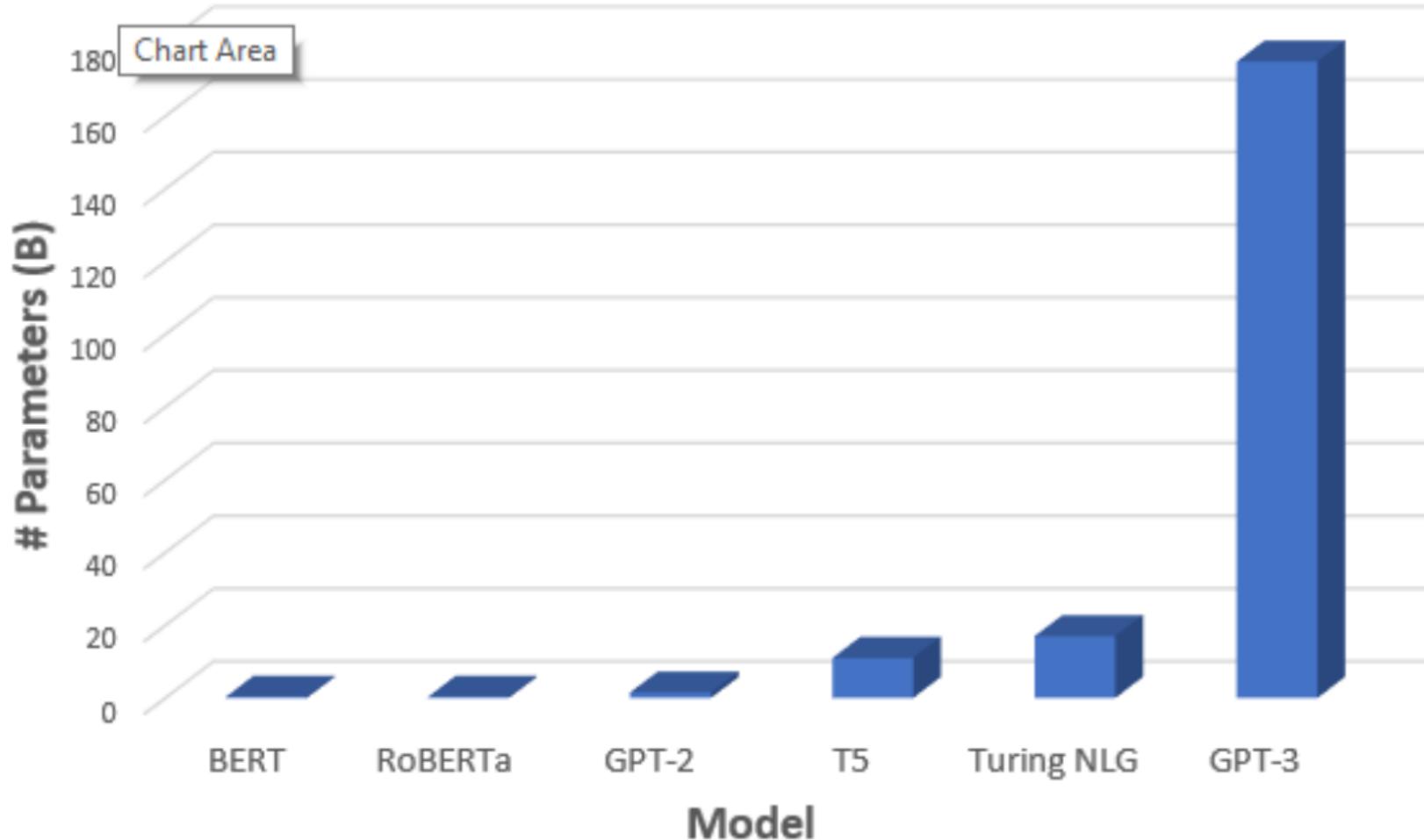


T5

TRANSFORMERS

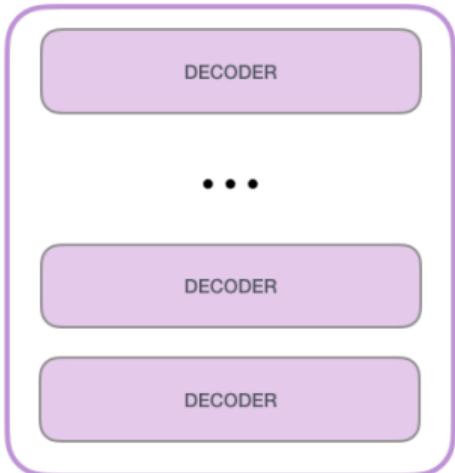




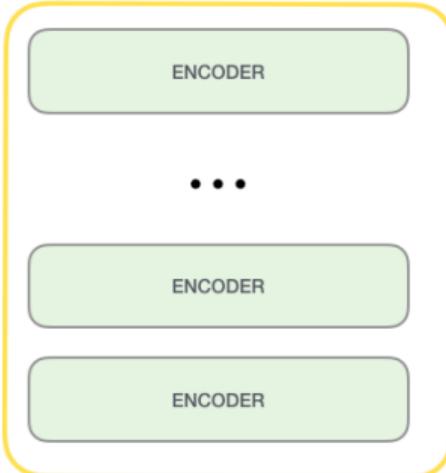




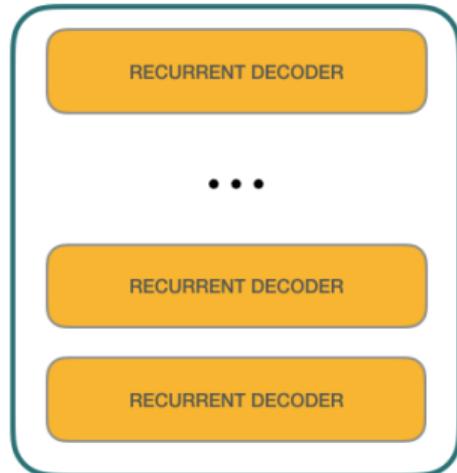
GPT-2



BERT



TRANSFORMER XL



Numbers of Parameters (in Millions)

20000

15000

10000

5000

0

AI2
ELMo
94

OpenAI
GPT
110

BERT
340

Ai2
Transformer
ELMo
465

OpenAI
GPT-2
1500

MT-DNN
330

XLM
665

UNIVERSITY of WASHINGTON
Grover
1500

Carnegie
Mellon
University
340

DistilBERT
66

RoBERTa
355

NVIDIA.
MegatronLM
8300

BART
400

XLM-R
550

Turing-NLG
17000

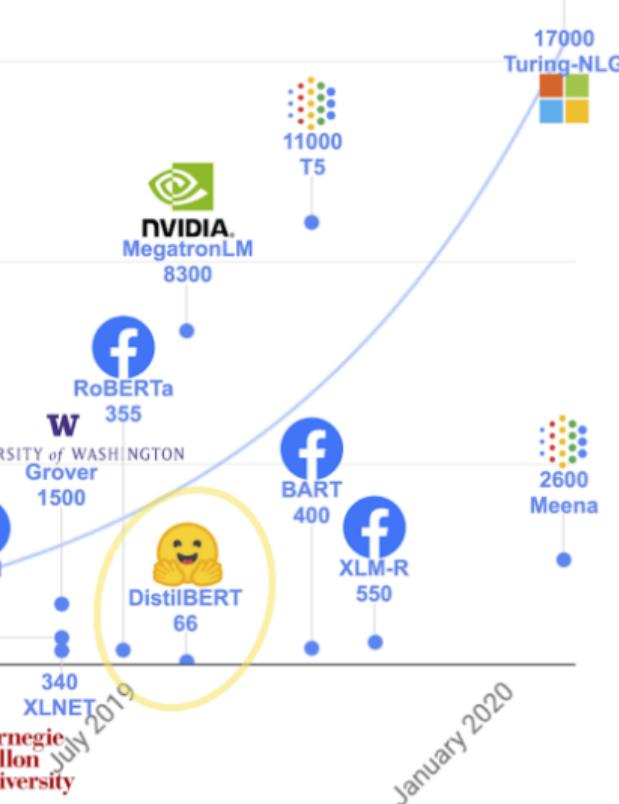
Meena
2600

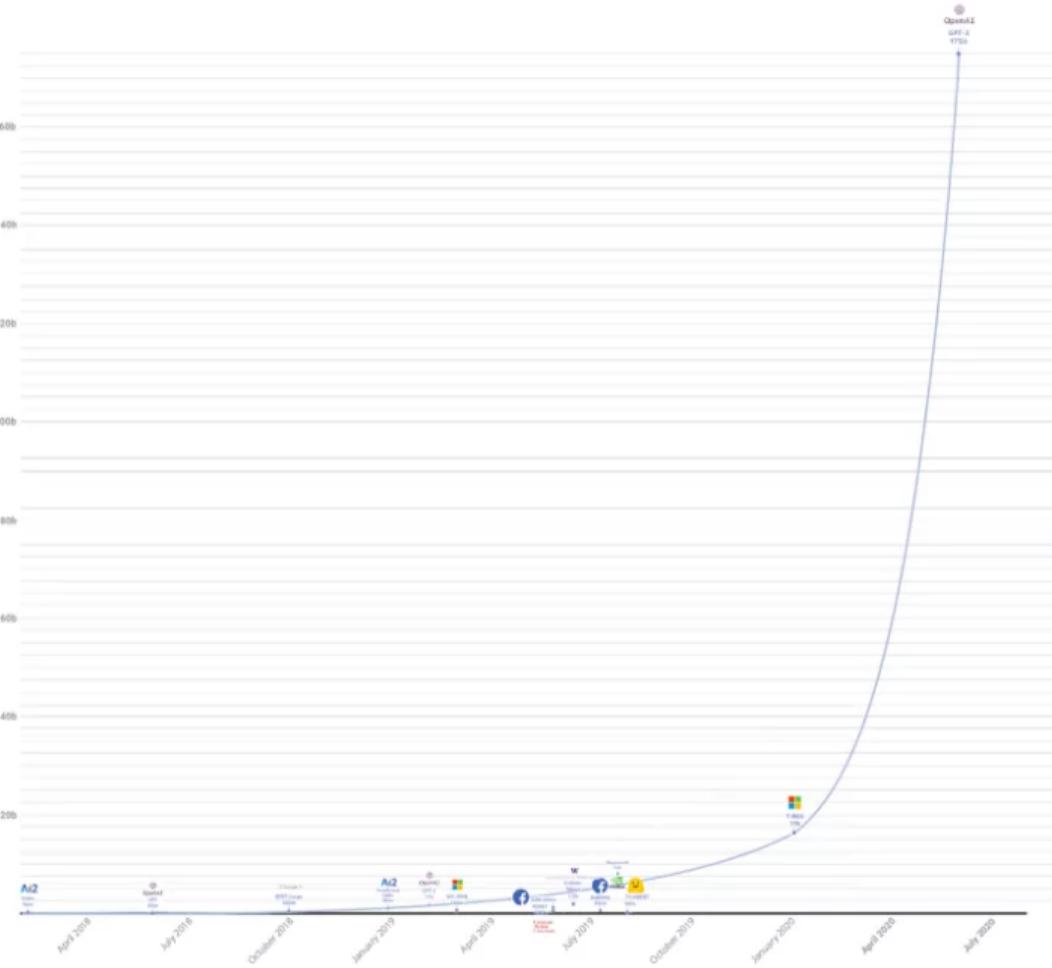
July 2018

January 2019

July 2019

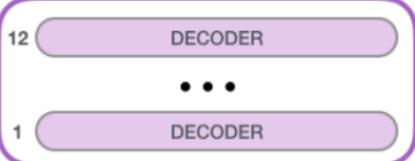
January 2020







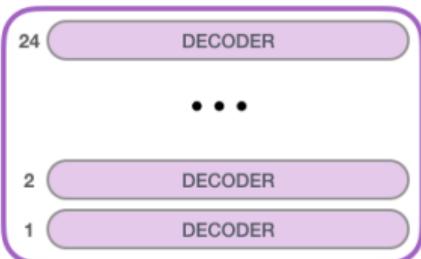
GPT-2
SMALL



Model Dimensionality: 768



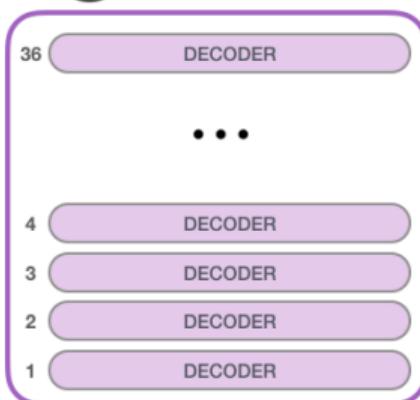
GPT-2
MEDIUM



Model Dimensionality: 1024



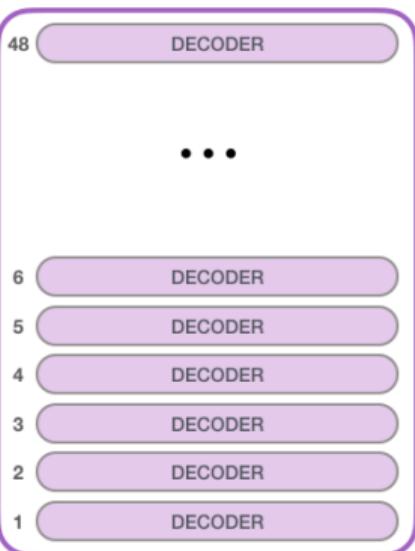
GPT-2
LARGE



Model Dimensionality: 1280



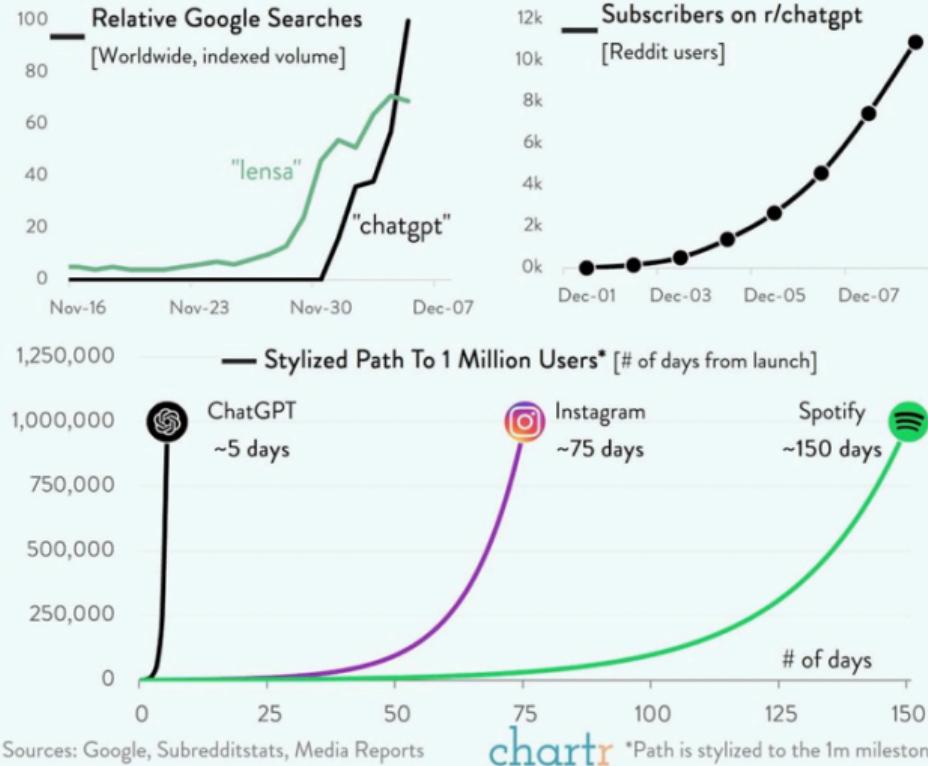
GPT-2
EXTRA
LARGE



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}

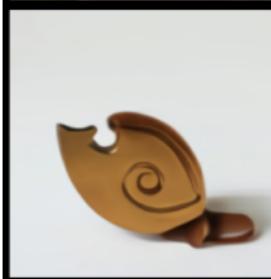
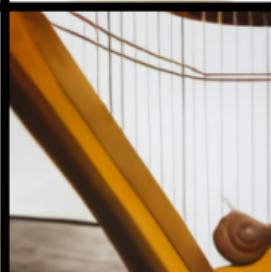
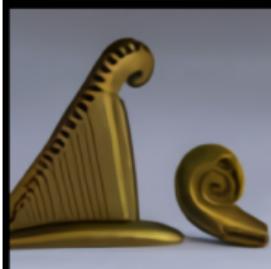
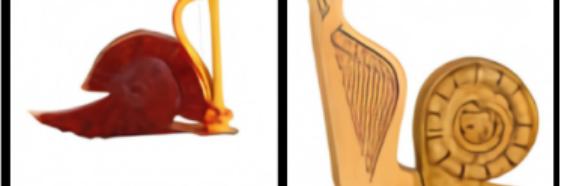
ChatGPT From OpenAI Is A Bot Taking The Tech World By Storm





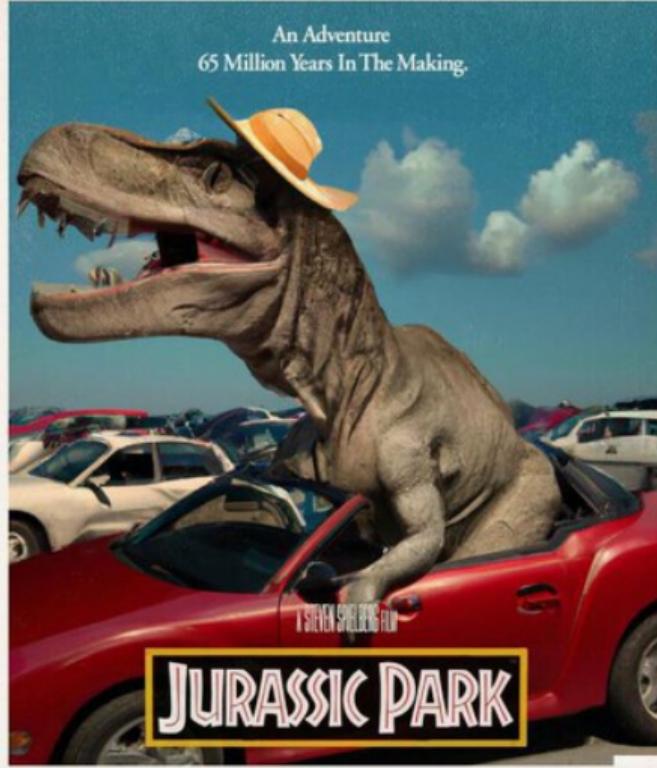
DALL-E









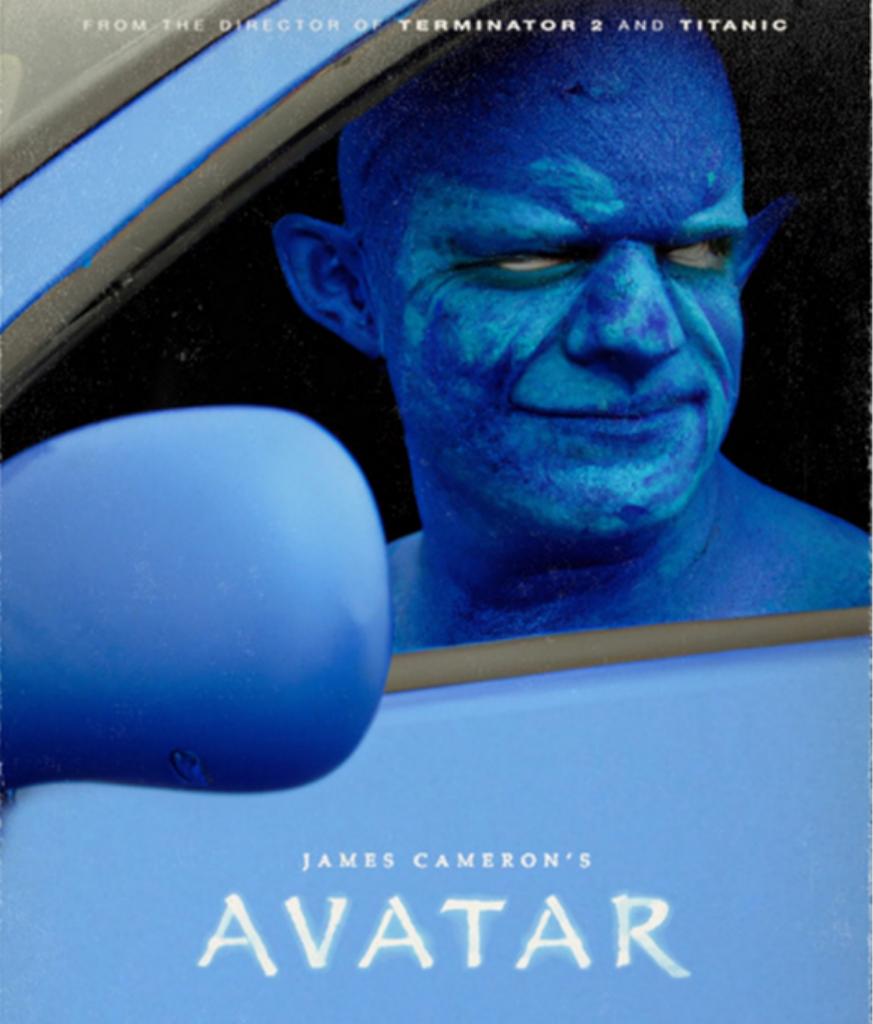


UNIVERSAL PICTURES PRESENTS AN AMBLIN ENTERTAINMENT PRODUCTION SAM NEILL LAURA DERN JEFF GOLDBLUM AND RICHARD ATTENBOROUGH "JURASSIC PARK" BOB PECK MARTIN FERERO BD WONG SAMUEL L. JACKSON WAYNE KNIGHT JOSEPH MAZZELLO ARIANA RICHARDS MUSIC BY STAN WINSTON DIRECTED BY DENNIS MUREN EDITED BY PHIL TIPPETT SPECIAL EFFECTS MICHAEL LANIER PROPS JOHN WILLIAMS FILM EDITOR MICHAEL KAHN PROPS RICK CARTER DIRECTED BY DEAN CUNDY STYLING MICHAEL CRUTCHLOW SCENIC DESIGN MICHAEL CRUTCHLOW AND DAVID KIEFER PROPS KATHLEEN KENNEDY AND GERALD R. MURKIN BASED ON THE MICHAEL CRUTCHLOW STORY SCREENPLAY BY MICHAEL CRUTCHLOW AND DAVID KIEFER PROPS KATHLEEN KENNEDY AND GERALD R. MURKIN

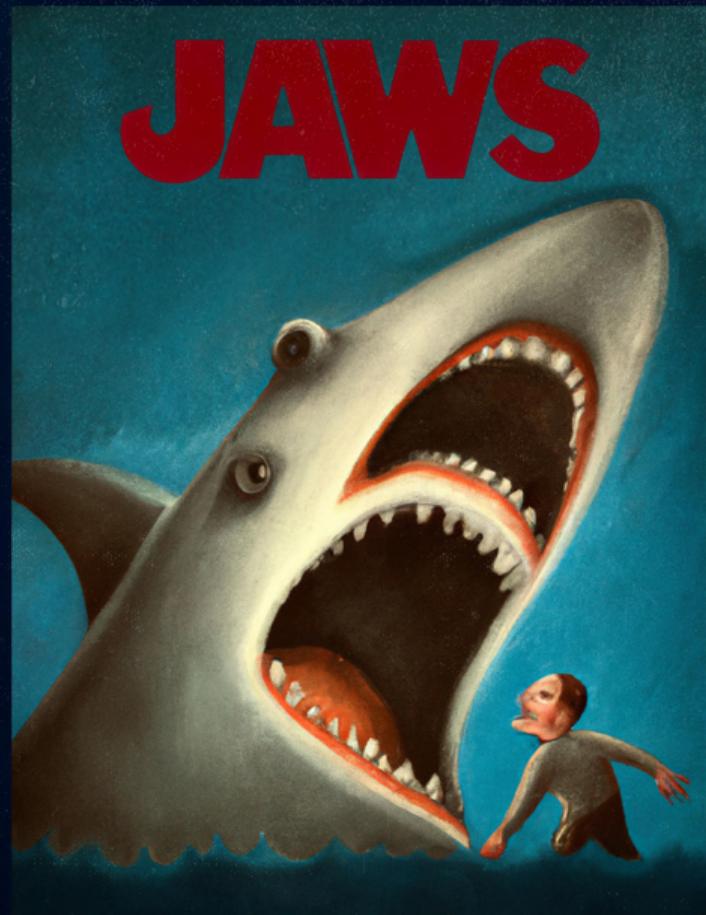


A STEVEN SPIELBERG FILM E.T. THE EXTRA-TERRESTRIAL
DEE WALLACE · PETER COYOTE · HENRY THOMAS AS ELLIOTT · MUSIC BY JOHN WILLIAMS
WRITTEN BY MELISSA MATHISON · PRODUCED BY STEVEN SPIELBERG & KATHLEEN KENNEDY
DIRECTED BY STEVEN SPIELBERG · A UNIVERSAL PICTURE
RATED PG-13 © 1982 UNIVERSAL CITY STUDIOS INC. MCMLXXXII AND TALES

FROM THE DIRECTOR OF TERMINATOR 2 AND TITANIC



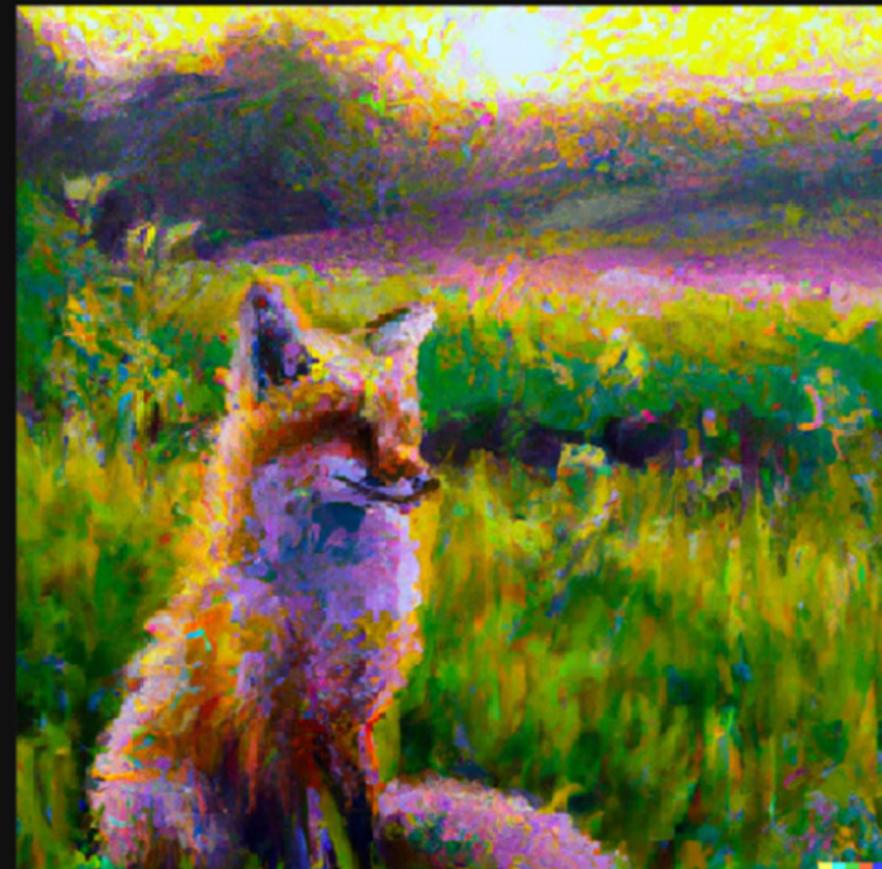
JAMES CAMERON'S
AVATAR



DALL·E 1



DALL·E 2



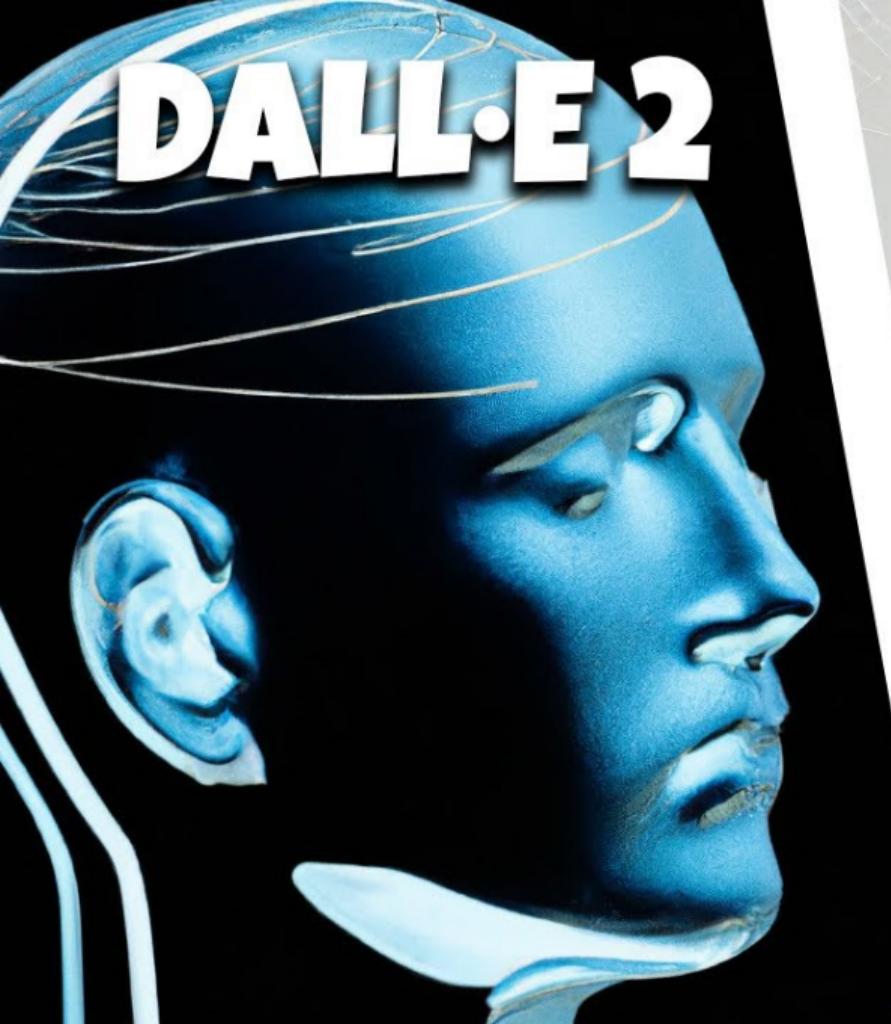


DALL-E 2 AI IMAGE GENERATORS



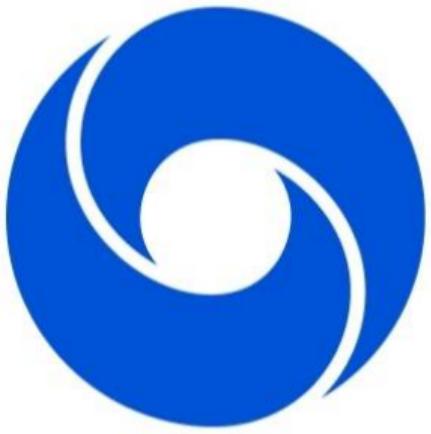


DALL·E 2

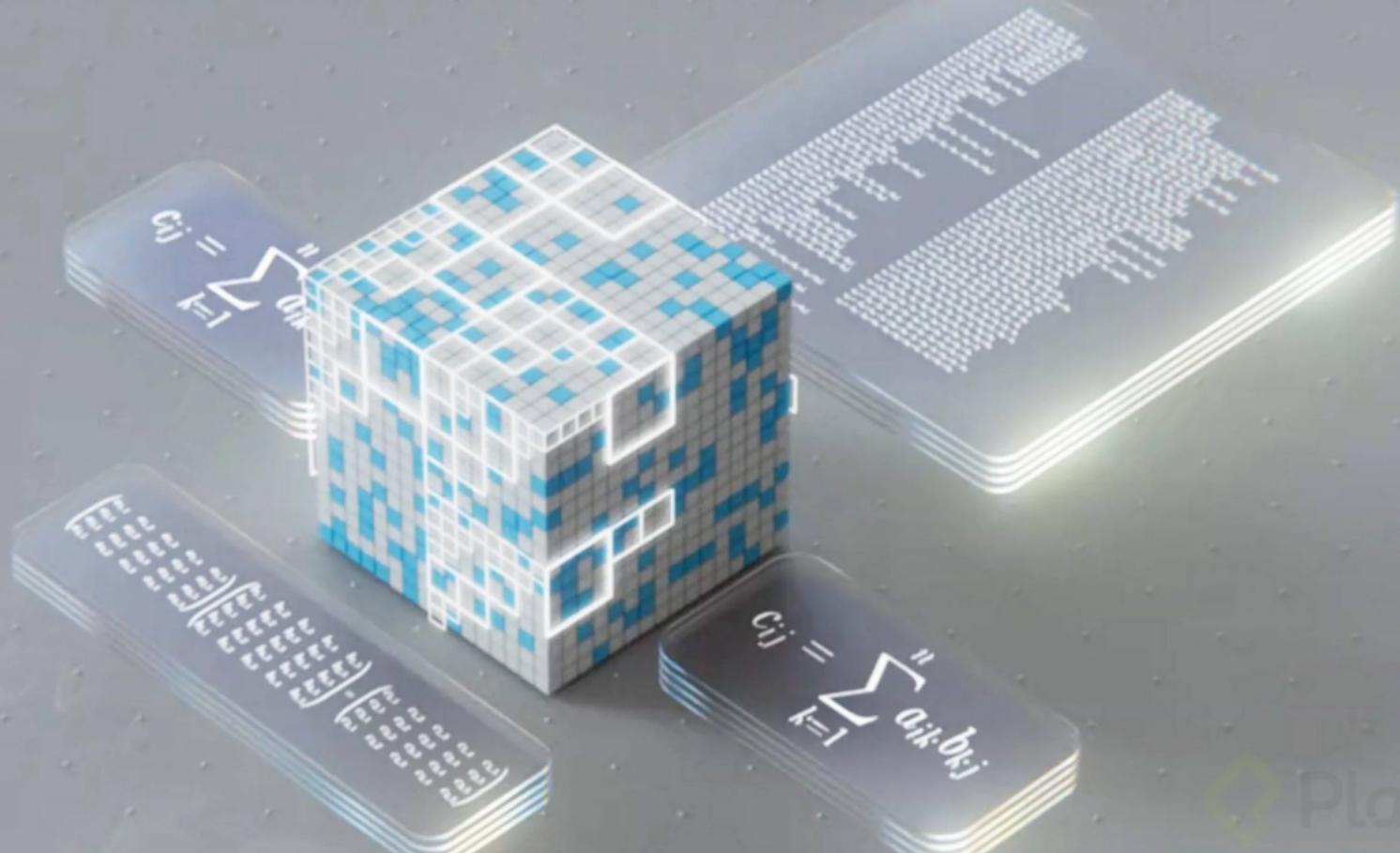


MIDJOURNEY v4





DeepMind



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

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

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