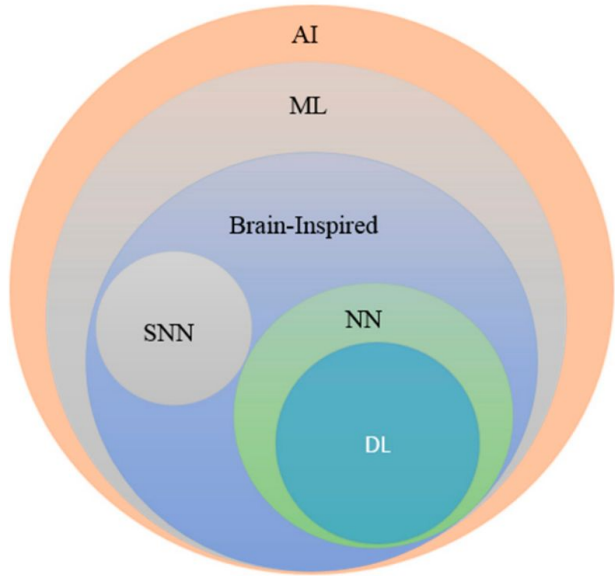


Temas selectos de Ciencia de Datos: Aprendizaje Profundo

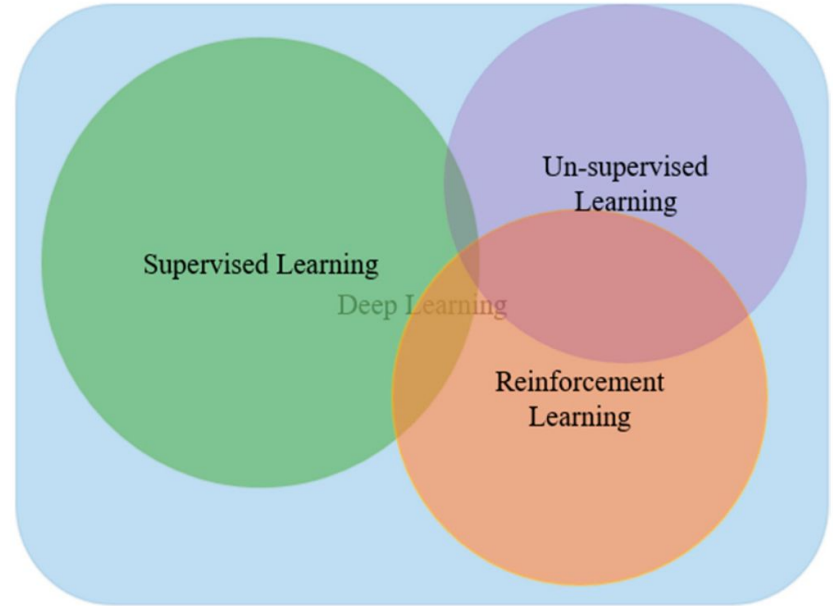
Presentación del Curso

Orlando Ramos Flores
Enero 2022

Taxonomía de la IA



AI: Artificial Intelligence. ML: Machine Learning.
NN: Neural Networks. DL: Deep Learning. SNN:
Spiking Neural Networks. Fuente [2]

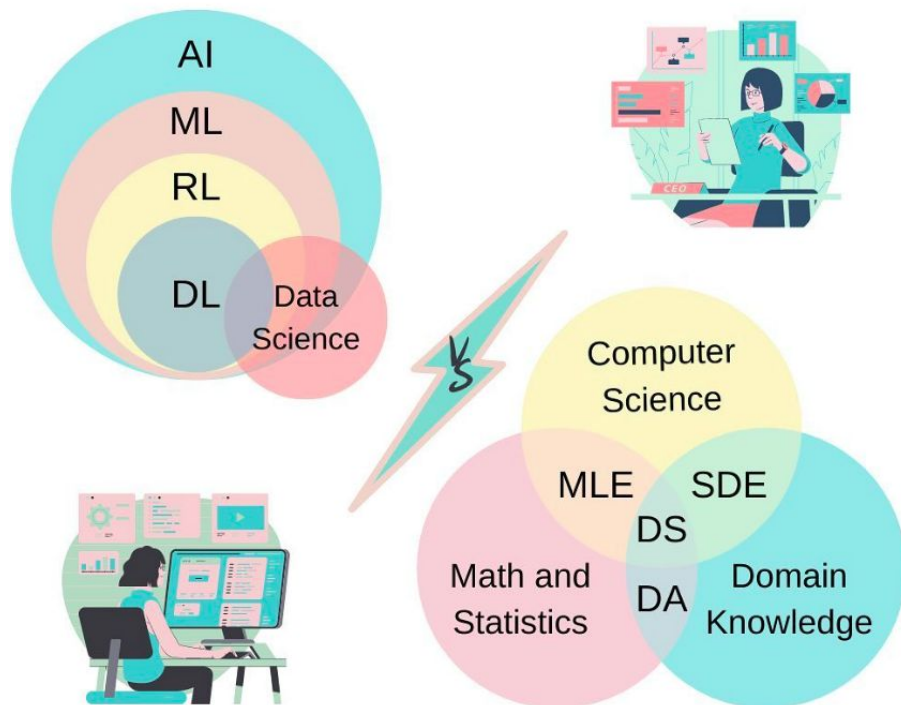


Categoría de enfoques de aprendizaje profundo.
Fuente [2]

Demystifying AI, ML, Data Science

Choose the career that suits YOU

@BigDataQueen



Glosario:

AI: Artificial Intelligence

ML: Machine Learning

RL: Representation Learning

DL: Deep Learning

DS: Data Scientist

DA: Data Analyst

MLE: Machine Learning Engineer


SDE: Software Development Engineer

Aprendizaje profundo

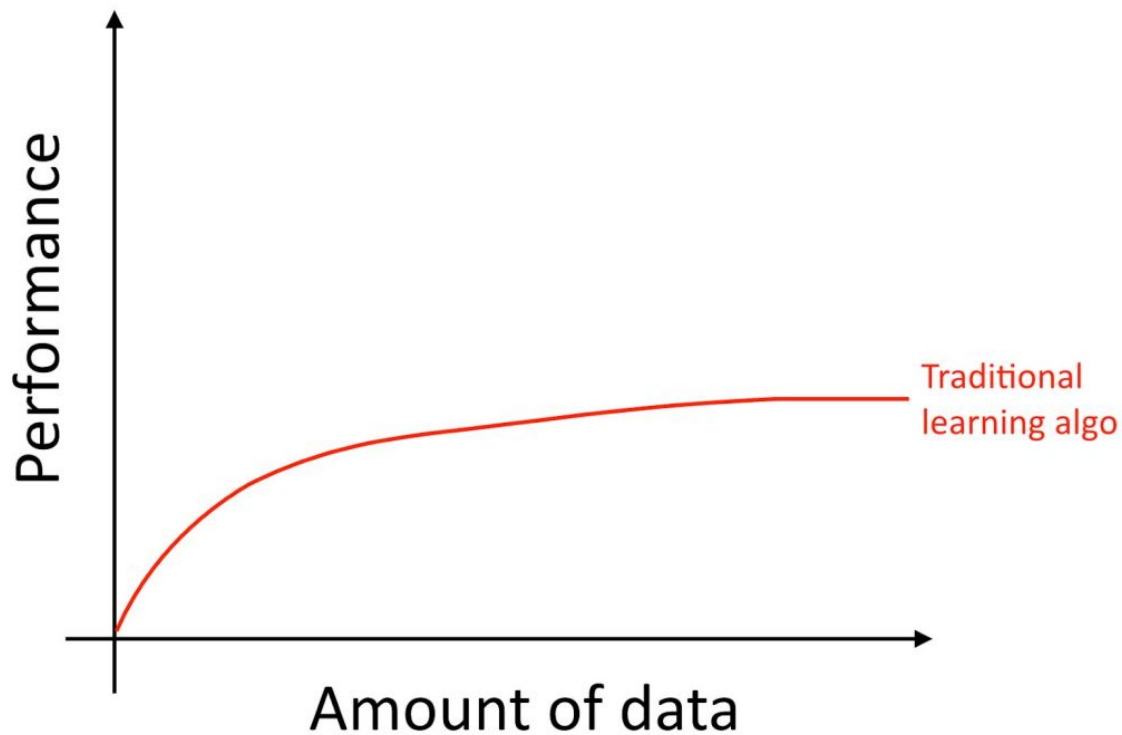
Frank Rosenblatt desarrolló y exploró todos los ingredientes básicos de los sistemas de aprendizaje profundo de hoy en día, y debería ser reconocido como el **Padre del Aprendizaje Profundo**, quizás junto con Hinton, LeCun y Bengio, quienes acaban de recibir el Premio Turing como los padres de la revolución del aprendizaje profundo [7].

¿Cuáles son los algoritmos desarrollados?

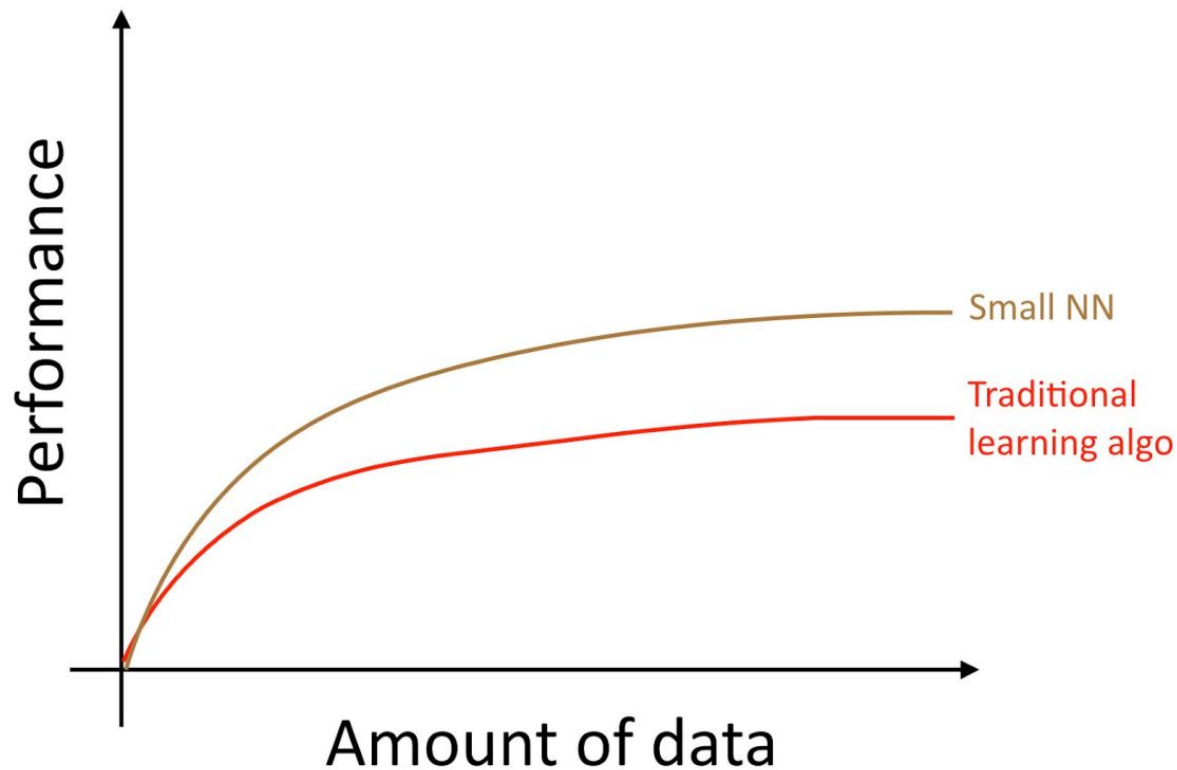
Las redes de avance profundo (Deep Feedforward Networks), también llamadas redes neuronales de avance (FNN: Feedforward Neural Networks), o perceptrones multicapa (MLP: Multilayer Perceptrons), son los modelos de aprendizaje profundo por excelencia [4]. Además, se mencionan las redes neuronales profundas (DNN: Deep Neural Networks) son redes neuronales que tienen múltiples capas entre las capas de entrada y salida [7].



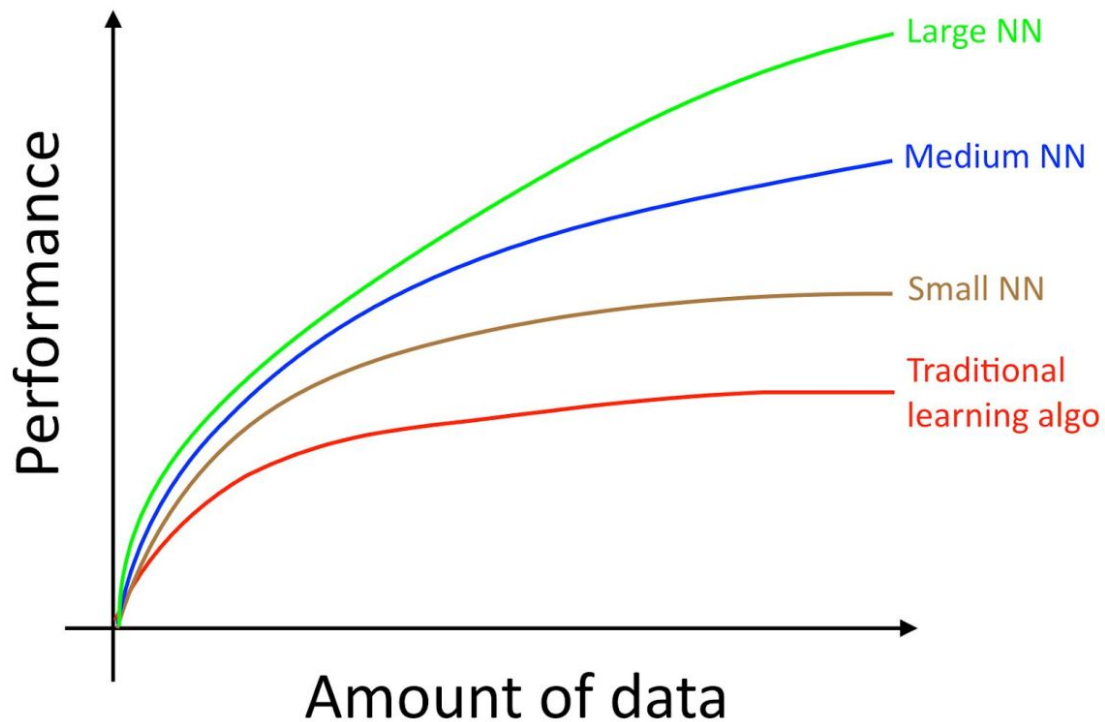
Progreso del Aprendizaje automático [5]



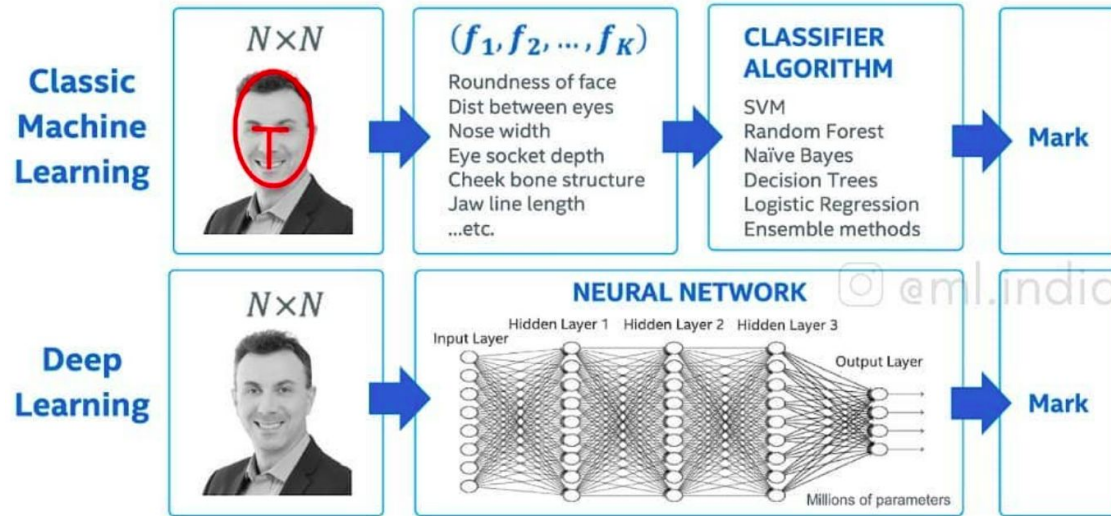
Progreso del Aprendizaje automático [5]



Progreso del Aprendizaje automático [5]

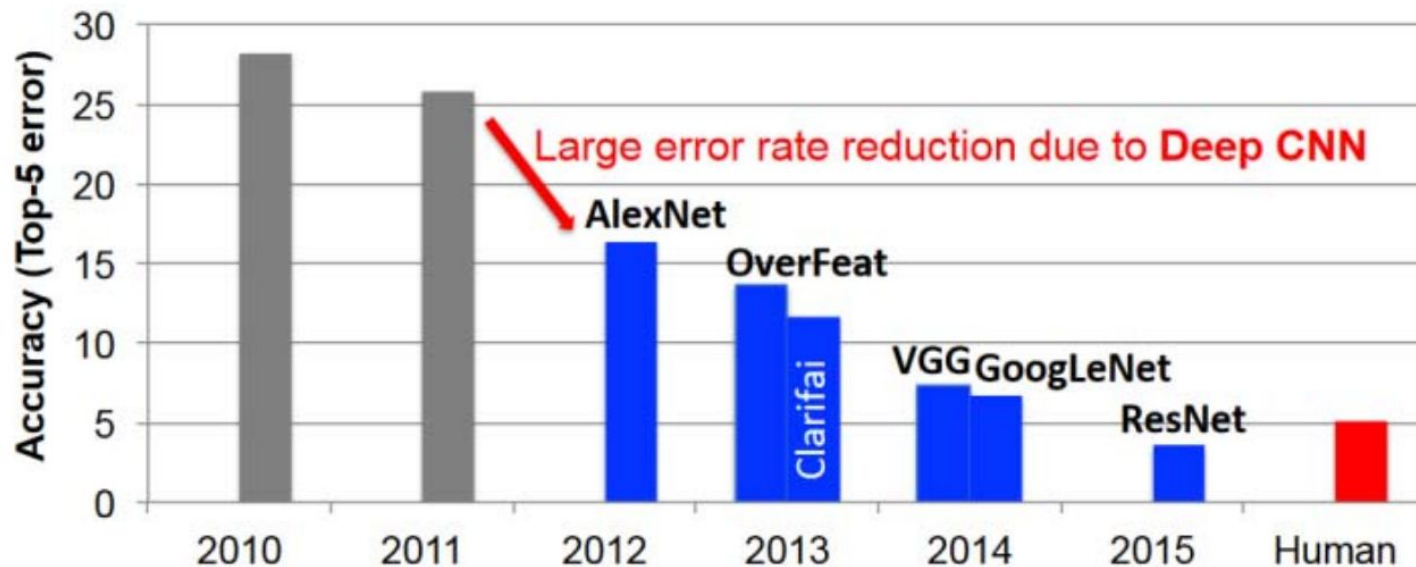


Aprendizaje profundo



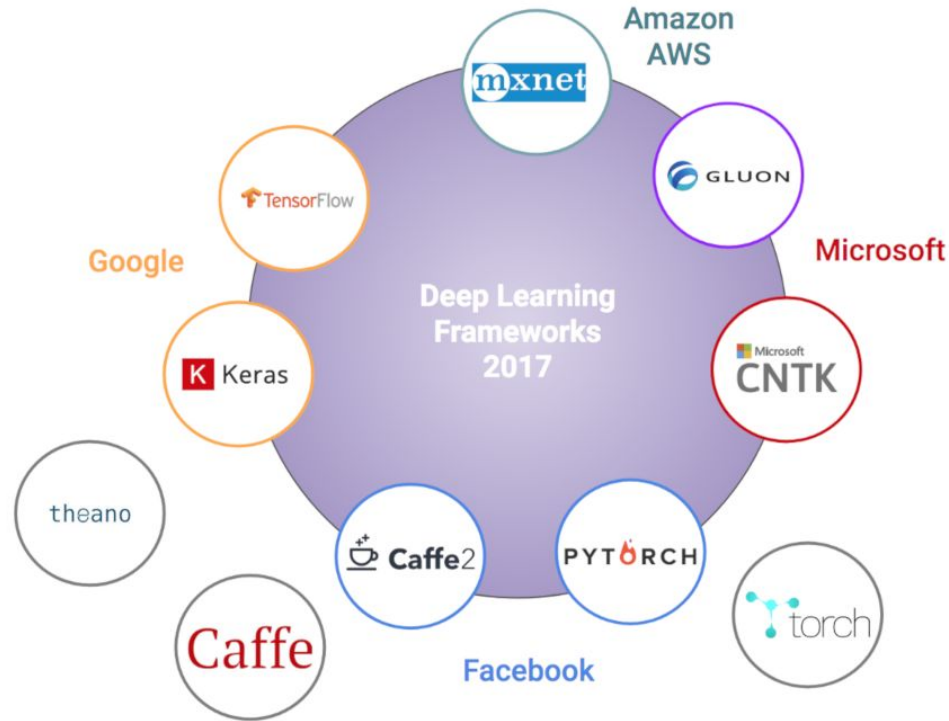
Source: <https://www.intel.com/>

Aprendizaje profundo



Una red convolucional ganó el Desafío de reconocimiento visual a gran escala de ImageNet (ILSVRC: ImageNet Large Scale Visual Recognition Challenge) por primera vez, y este desafío ahora lo ganan consistentemente las redes profundas [7].

Frameworks de Deep Learning [1]



Generalización

Durante siglos, los científicos, los políticos, los actuarios y los vendedores han explotado el hecho empírico de que los resultados desconocidos, ya sean futuros o no observados, a menudo rastrean regularidades encontradas en observaciones pasadas.

Llamamos a esta idea generalización: encontrar reglas consistentes con los datos disponibles que se aplican a instancias que aún tenemos que encontrar [8].



Consideraciones de energía y política para el aprendizaje profundo en NLP [3]

Consumption	CO ₂ e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Table 1: Estimated CO₂ emissions from training common NLP models, compared to familiar consumption.¹

Consumer	Renew.	Gas	Coal	Nuc.
China	22%	3%	65%	4%
Germany	40%	7%	38%	13%
United States	17%	35%	27%	19%
Amazon-AWS	17%	24%	30%	26%
Google	56%	14%	15%	10%
Microsoft	32%	23%	31%	10%

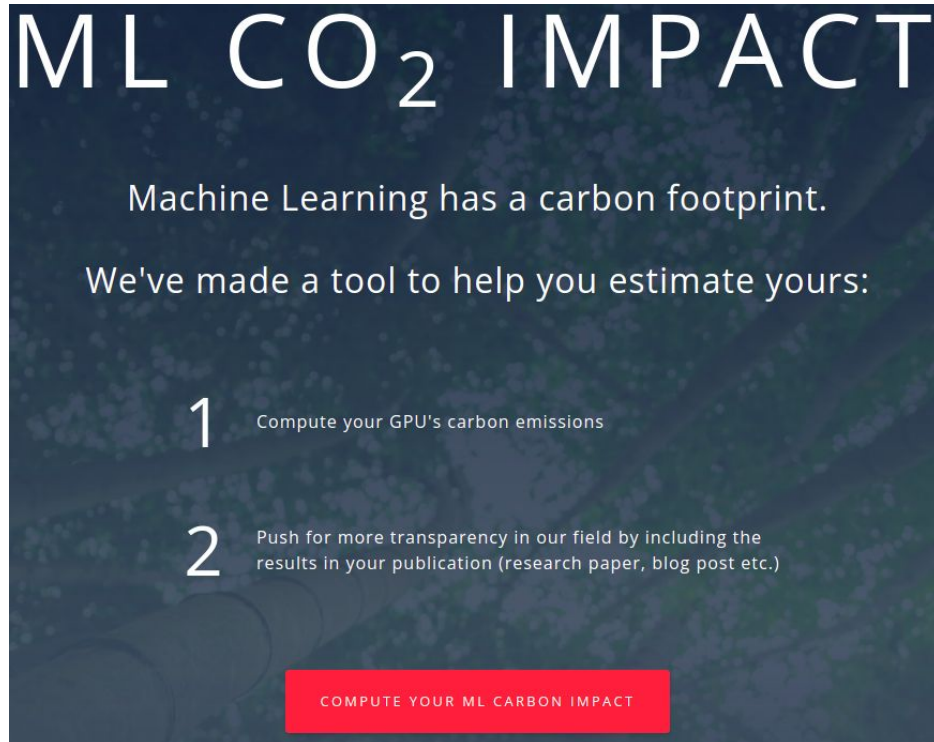
Table 2: Percent energy sourced from: Renewable (e.g. hydro, solar, wind), natural gas, coal and nuclear for the top 3 cloud compute providers (Cook et al., 2017), compared to the United States,⁴ China⁵ and Germany (Burger, 2019).

Consideraciones de energía y política para el aprendizaje profundo en NLP [3]

Model	Hardware	Power (W)	Hours	kWh·PUE	CO ₂ e	Cloud compute cost
Transformer _{base}	P100x8	1415.78	12	27	26	\$41–\$140
Transformer _{big}	P100x8	1515.43	84	201	192	\$289–\$981
ELMo	P100x3	517.66	336	275	262	\$433–\$1472
BERT _{base}	V100x64	12,041.51	79	1507	1438	\$3751–\$12,571
BERT _{base}	TPUv2x16	—	96	—	—	\$2074–\$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	\$942,973–\$3,201,722
NAS	TPUv2x1	—	32,623	—	—	\$44,055–\$146,848
GPT-2	TPUv3x32	—	168	—	—	\$12,902–\$43,008

Table 3: Estimated cost of training a model in terms of CO₂ emissions (lbs) and cloud compute cost (USD).⁷ Power and carbon footprint are omitted for TPUs due to lack of public information on power draw for this hardware.

Calculadora de emisiones de aprendizaje automático [6]



<https://mlco2.github.io/impact/>

Temario I

- Redes Densas
 - Introducción
 - Neurona artificial (Perceptrón)
 - Red Neuronal Multicapa
- Redes Convolucionales
 - Red Convolucional para imágenes
 - Red Convolucional para texto
- Redes Recurrentes
 - Memoria a corto y largo plazo (LSTM)
 - Memoria Bidireccional a corto y largo plazo (Bi-LSTM)
 - Unidad recurrente cerrada (GRU)



Temario II

- Redes basadas en atención
 - Codificador y Decodificador
 - Aprendizaje de secuencia a secuencia
 - Atención de múltiples cabezas
 - Autoatención
 - Transformers
 - * [Hugging Face](#)
- Estrategias de entrenamiento
 - Configuración de conjuntos de desarrollo y prueba
 - Regularización
 - Optimización para el entrenamiento
 - Sesgo y varianza
 - Curvas de aprendizaje
 - Estrategias de optimización



Temario III

- Redes Generativas
 - Redes adversarias generativas (GAN)
- * Modelos de lenguaje y aprendizaje de representaciones no supervisado
 - ELMo
 - BERT
 - GPT
 - T5



Información del curso

Martes y Jueves

10:00 AM a 12:00 PM

Credenciales de Zoom:

<https://cuaieed-unam.zoom.us/j/2637875056?pwd=Uk1vTmFZeGVNTHpBWGY5eS9mUIBYUT09>

ID de reunión: 263 787 5056

Código de acceso: 865146

Contacto:

Orlando Ramos Flores

orlando.ramos@aries.iimas.unam.mx




Información adicional del curso


Repositorio del curso:

<https://github.com/orlandxrf/curso-dl>

- data
- notebooks
- slides

Usar de forma **local** o con **Google Colab** 

Lenguaje de programación: **Python** 

Biblioteca de Python para el curso: **Pytorch** 

Criterios de evaluación

- 30% Tareas
 - Trabajo para practicar lo aprendido
- 30% Exámenes (2 parciales)
 - Mini Proyecto sobre los temas vistos
 - Definición competencias en Kaggle (<https://www.kaggle.com/competitions>)
- 40% Proyecto final
 - Reproducir y replicar resultados de artículos o tareas compartidas
 - Aplicar a otras tareas/condiciones realizando un nuevo análisis o modificando la arquitectura
 - Establecer un proyecto propio, empleando lo visto en el curso



Referencias

1. den Bakker, Indra. 2017. "Battle of the Deep Learning frameworks – Part I: 2017, even more frameworks and interfaces." Towards Data Science, December 2021. Accessed 2021-12-20.
2. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3), 292.
3. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243*.
4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press. URL: <http://www.deeplearningbook.org>
5. Andrew, N. (2020). Khát Khao Học Máy (Machine Learning Yearning).
6. Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.
7. Tappert, C. C. (2019, December). Who Is the Father of Deep Learning?. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 343-348). IEEE.
8. Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3), 107-115.

