



PRESENTACIÓN 2: REDES NEURONALES

Machine Learning II

Jose Alexander Fuentes Montoya Ph.D(c)

SEGUNDA PARTE

INTRODUCCIÓN AL DEEP LEARNING

NEURONAS

El estudio de cómo funciona el cerebro ha fascinado a los científicos durante mucho tiempo. Esta fascinación aumentó con la llegada de la histología, que reveló cómo se organizan las neuronas.

La doctrina de la neurona afirma que el sistema nervioso está compuesto de **neuronas independientes**. Antes, se creía ampliamente que el sistema nervioso consistía en una red continua única, teoría propuesta por Joseph von Gerlach y defendida por Camillo Golgi. El término “neurona” fue propuesto por Heinrich Wilhelm Gottfried Waldeyer-Hartz hacia 1891.

Ambos, Golgi y Cajal, (Figura 1) recibieron el Premio Nobel en 1911 por su trabajo. Posteriormente, la microscopía electrónica demostró que las neuronas son, de hecho, células independientes. Estas neuronas **inspiraron las Redes Neuronales**. La caricatura mostrada en la Figura 2 se elaboró usando Imagen, aplicación que se basa en **Redes Neuronales**. La línea temporal se muestra en la Figura 3.

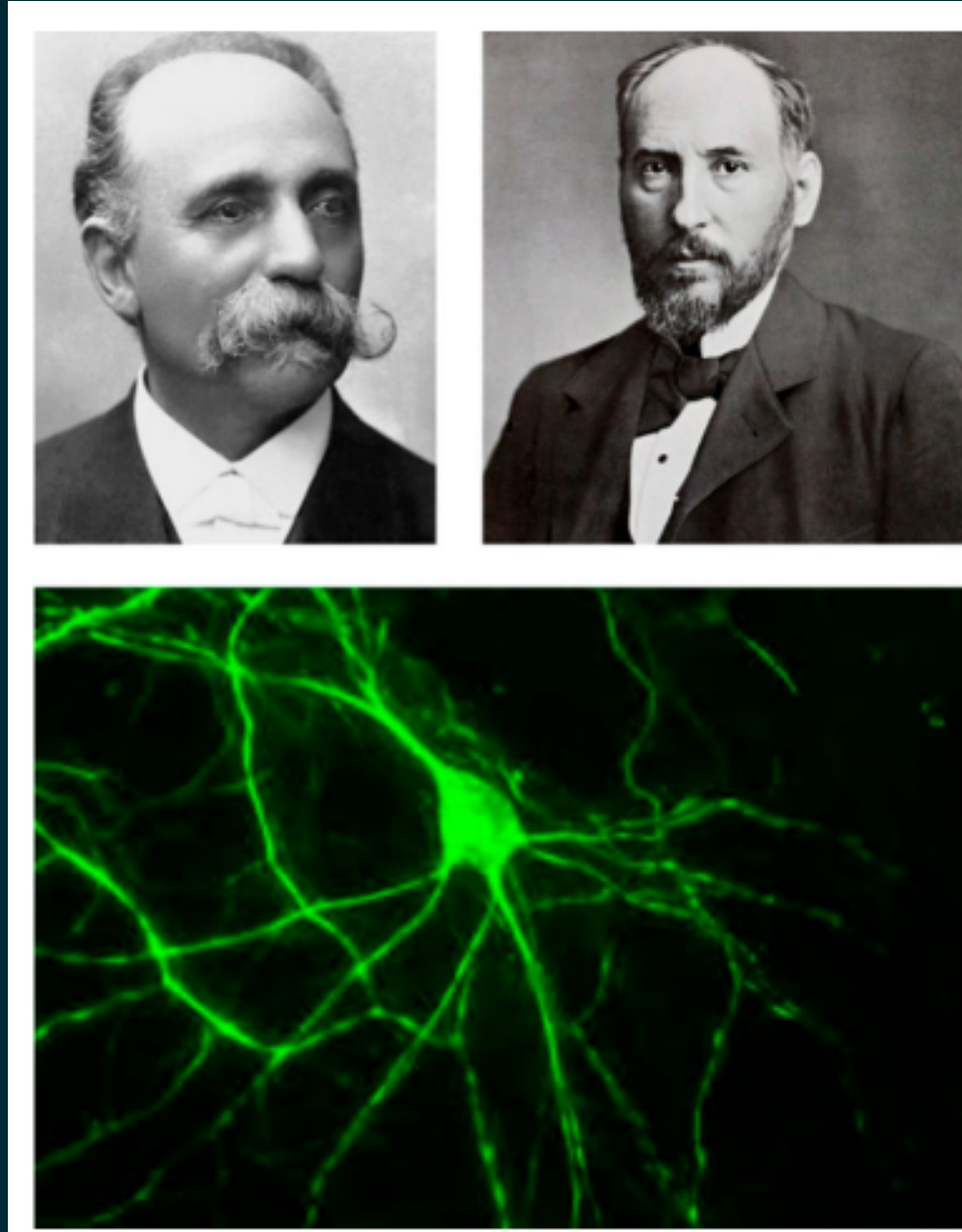


Figura 1. Redes neuronales

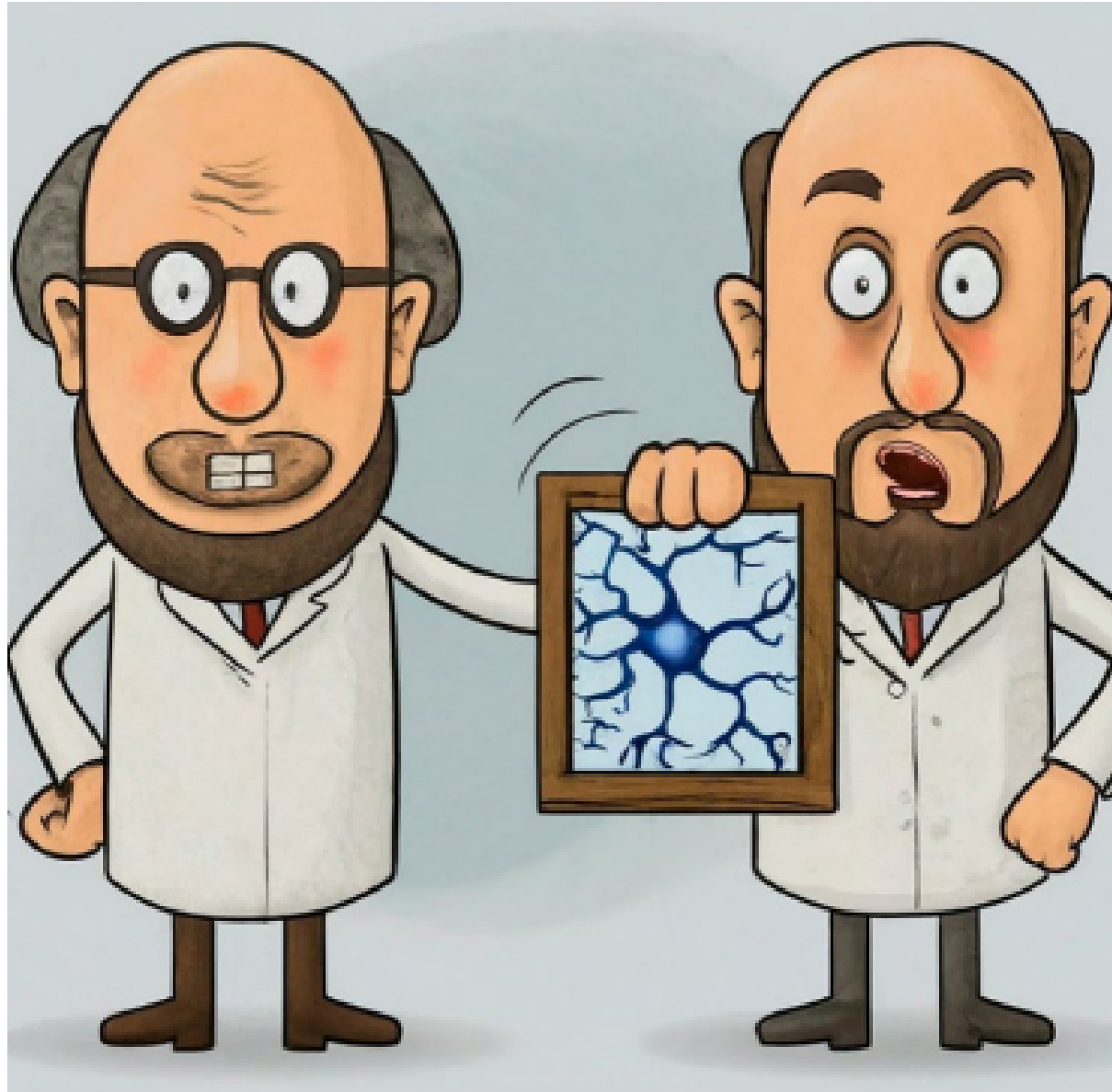


Figura 2. Caricatura generad por IA

Neuron

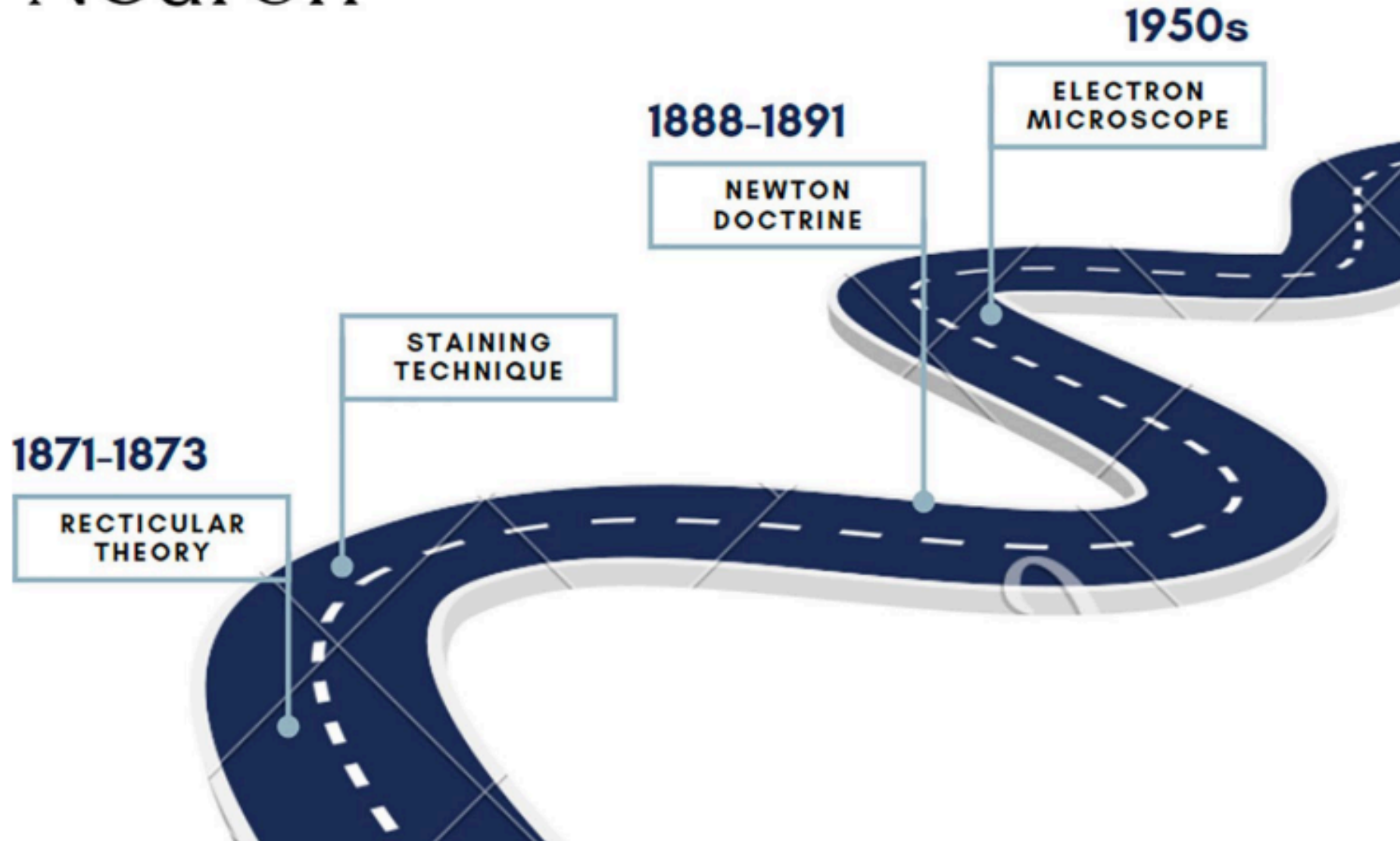


Figura 3. Línea de tiempo del Neuron

DEL PERCEPTRÓN AL INVIERNO DE LA IA

Los modelos aca tratados son **Redes Neuronales Profundas (DNNs)**, basadas en Redes Neuronales que, a su vez, se inspiran en la neurona biológica.

El primer modelo computacional inspirado en una neurona fue el modelo McCulloch–Pitts, propuesto por un neurólogo (McCulloch) y un lógico (Pitts). El modelo tenía **entradas binarias** $x_1, x_2, x_3, \dots, x_n$ $x_i \in \{0, 1\}$, salida binaria $y \in \{0, 1\}$ y una unidad de umbral (*thresholding*).

Estos modelos podían implementar *logic gates* y, por tanto, se estableció que podían implementar una “máquina lógica”.

Más tarde, Frank Rosenblatt propuso pesos e inputs continuos, lo que mejoró notablemente el poder de los perceptrones. Entonces podían usarse para **clasificación lineal y regresión**. La gente dijo que podrían

| (“gobernar el mundo”).

En particular, demostraron que los perceptrones **no podían resolver el problema XOR**. Esto condujo a un declive del interés en las Redes Neuronales desde 1969 hasta mediados de los 80 (a lo que a menudo se llama

| (“Invierno de la IA”).

En 1986, Rumelhart et al. propusieron el **algoritmo de retropropagación (backpropagation)**, que permitió entrenar un *Multi-Layer Perceptron (MLP)*. Esto posibilitó redes de múltiples capas que **sí resolvían** el XOR y mejoraron considerablemente en diversas tareas de aprendizaje supervisado. En 1991 se introdujo la idea de preentrenar redes, sentando las bases de muchos trabajos posteriores a 2006.

NEURAL NETWORK TIMELINE

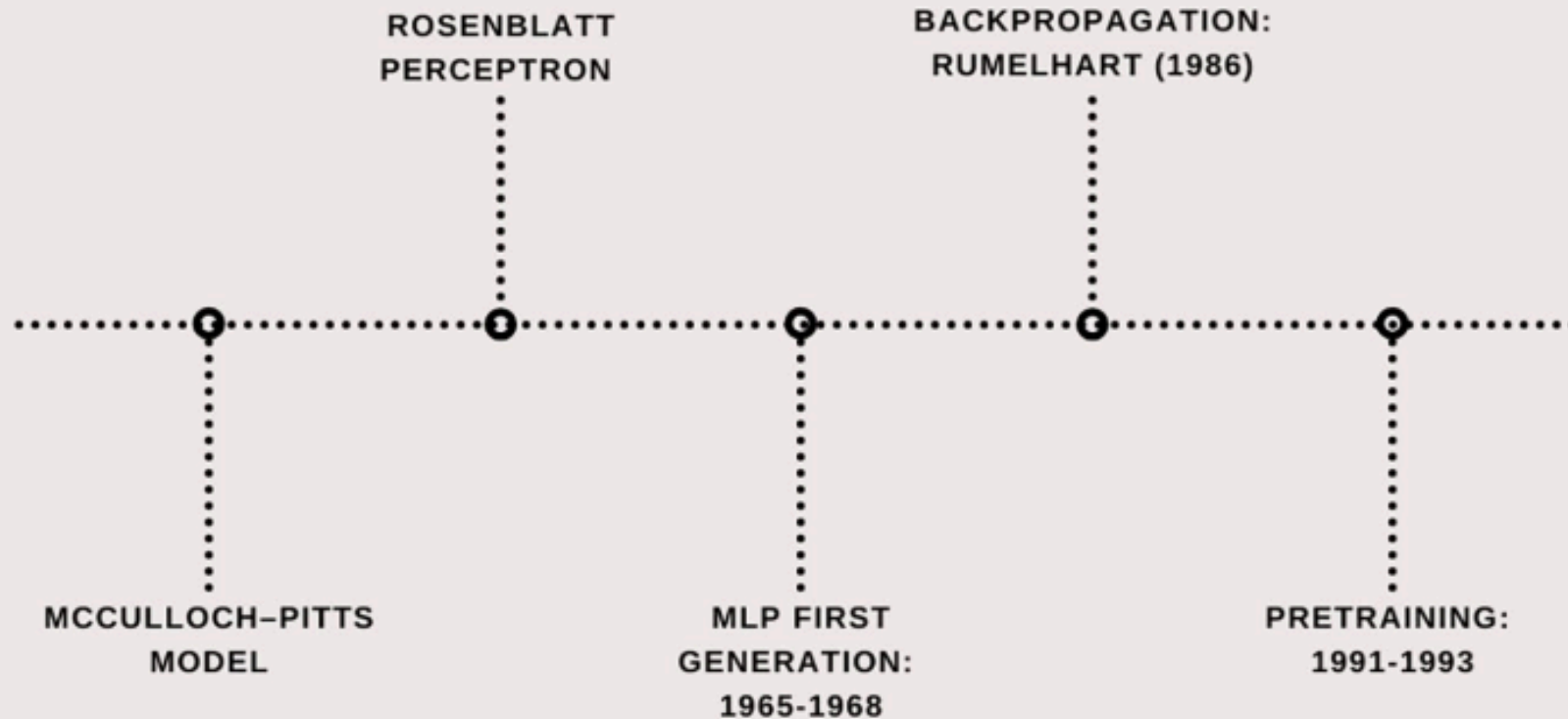


Figura 4. Línea temporal: Redes Neuronales

PROCESAMIENTO DE IMÁGENES Y REDES NEURONALES CONVOLUCIONALES

Para tareas relacionadas con imágenes, se necesitaba una red capaz de **inferir correlaciones espaciales**.

Curiosamente, **los gatos** ayudaron a la comunidad científica a desarrollar estas redes.

Un experimento con gatos demostró que solamente algunas partes del cerebro se activan en respuesta a cierto estímulo (Figura 5). Dicho experimento fue realizado por Hubel y Wiesel, quienes demostraron lo siguiente:

“Una Neurona se activa solo en respuesta a un estímulo particular en una región particular.”

El desarrollo de las **Redes Neuronales Convolucionales (CNN)**, basadas en este concepto, data de la década de 1990. Una CNN llamada **LeNet** se propuso para identificar dígitos manuscritos. Ese trabajo también nos dio el famoso **conjunto de datos MNIST**, que luego muchos científicos usaron para probar sus modelos.



Figura 5. Solo algunas partes del cerebro de un gato se activan al ver cierta imagen

Los avances en análisis de imágenes se impulsaron con la competición **ImageNet** (1000 clases y numerosas imágenes). Según su sitio oficial:

“ImageNet es una base de datos de imágenes organizada según la jerarquía de WordNet. Cada nodo de la jerarquía está representado por cientos o miles de imágenes. El proyecto ha sido clave para el avance de la visión por computadora y la investigación en deep learning. Los datos están disponibles gratis para uso de investigación no comercial.” [3]

Los modelos que ganaron o tuvieron alto rendimiento en este concurso se volvieron relevantes en el área. Algunos de ellos fueron:

- **AlexNet**, con 8 capas
- **ZFNet**, también 8 capas pero mejor tasa de error que AlexNet
- **VGG-Net**, con 19 capas, mejor tasa de error
- **GoogLeNet**
- **ResNet**

Uno de los primeros logros fue el de **reconocimiento de escritura a mano**. Graves et al. superaron en 2009 el estado del arte en reconocimiento de escritura árabe. Cireşan et al. crearon un referente en MNIST. El año siguiente introdujo un reconocedor de patrones para el sistema IJCNN de reconocimiento de señales de tráfico.

En 2016 se dieron avances notables en reconocimiento de voz (Speech Recognition), con una mejora de 16% respecto al estado del arte en ciertos conjuntos de datos.

THE ADVENT OF DEEP LEARNING

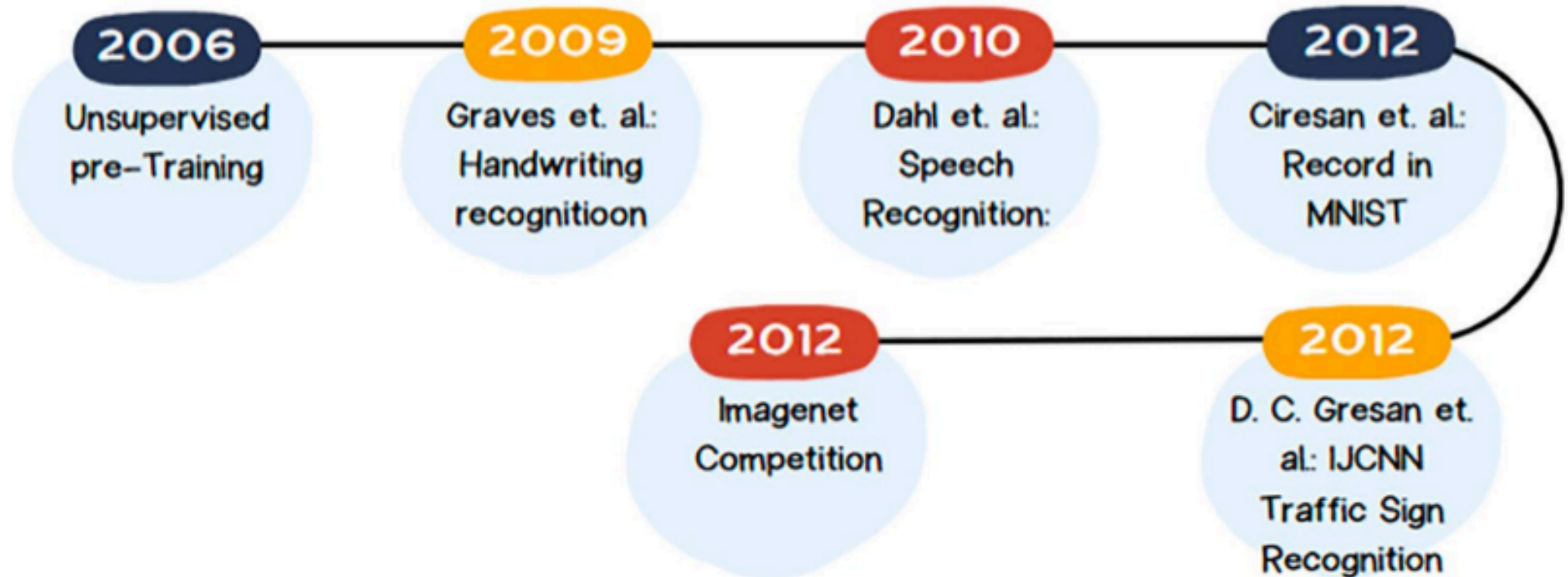


Figura 6. Auge del Deep Learning

¿QUÉ HAY DE NUEVO?

La aparición de mejores optimizadores condujo a convergencias más rápidas y mejores precisiones. Desde el **gradient descent** tradicional, pasando por Nesterov (1983), AdaGrad (2011), RMSprop (2012) y Adam (2015), se ha ido facilitando el entrenamiento de redes profundas (Figura 2-5). De hecho, aparecieron otros algoritmos como Eve y Beyond Adam.

Asimismo, surgieron nuevas funciones de activación como **tanh** (1991), **ReLU** (2010), **Leaky ReLU** (2013) y **SIREN** (2020). Además, el progreso del hardware (GPUs, TPUs, etc.) también impulsó el desarrollo de Deep Learning.

Optimization algorithms

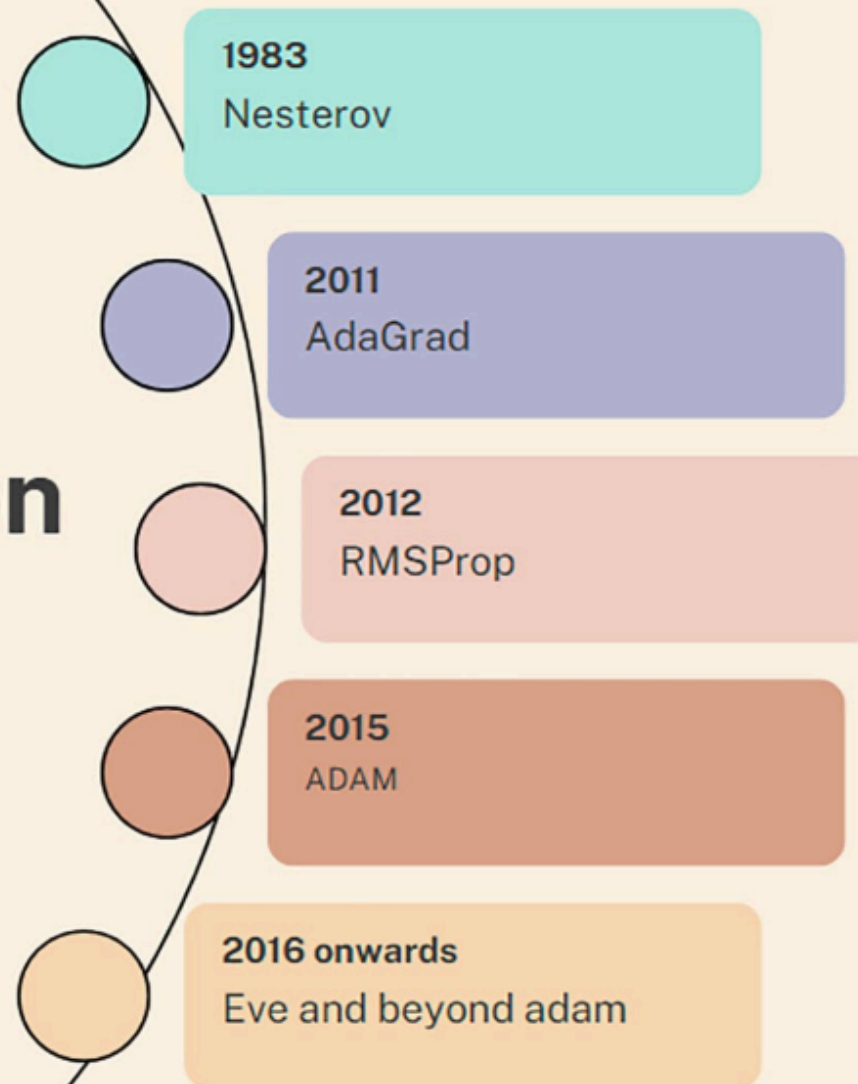


Figura 7. Algoritmos de optimización

SECUENCIAS

Aunque las redes totalmente conectadas ayudaron a resolver muchos problemas y las CNN resolvieron otros relacionados con imágenes, **quedaban pendientes los relacionados con secuencias** (texto, voz, series temporales, etc.). En esos casos, la relación entre diferentes pasos de la secuencia es clave.

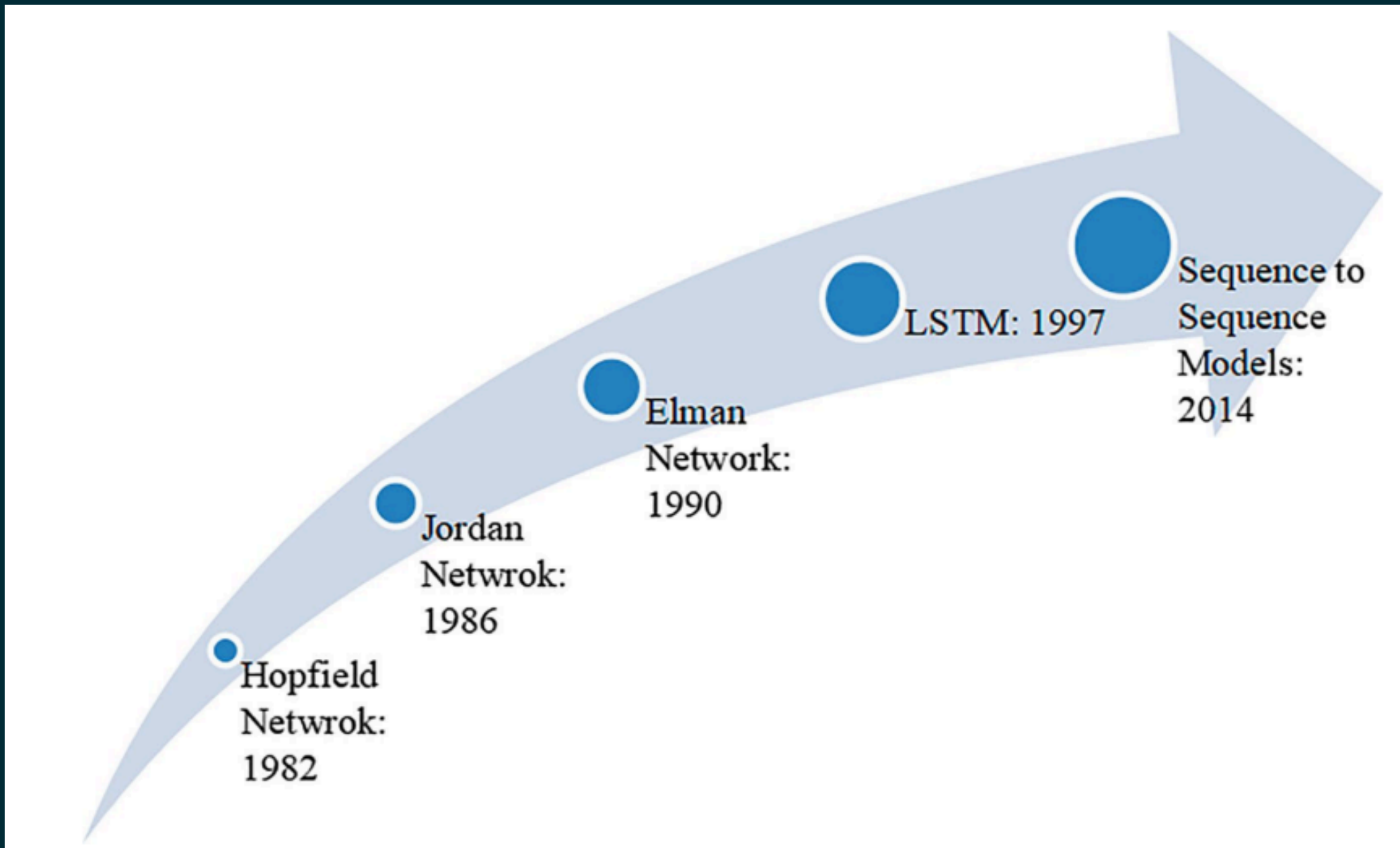


Figura 8. Línea temporal: modelos de secuencia

La **Red de Hopfield** (1982) modeló una “memoria asociativa por contenido”. La **Red de Jordan** planteó la idea de tomar la salida de un estado como entrada del siguiente. Del mismo modo, la **Red de Elman** propuso que el estado oculto de una red alimentara el estado oculto de la siguiente.

Las redes recursivas así desarrolladas sufrían problemas como el **desvanecimiento del gradiente (vanishing gradient)**. Modelos como **GRU** y **LSTM** abordaron este inconveniente.

LA DEFINICIÓN

Para tareas supervisadas y no supervisadas, se requieren:

- 1) Preprocesamiento,
- 2) Extracción de características y
- 3) Selección de características.

La extracción de características suele ser específica de la modalidad (imágenes, texto, sonido). Hay numerosos métodos y no es sencillo seleccionar el mejor. También la selección de características puede hacerse con métodos *filtro* o *wrapper*.

Los métodos de Deep Learning extraen y eligen automáticamente las características importantes, sin necesidad de especificar manualmente qué usar.

Además, **Deep Learning** suele dar mejor desempeño si hay suficientes datos. Emplea optimizadores de última generación y aprovecha la aceleración por hardware. Formalmente, se ha definido así:

“Los métodos de deep learning son métodos de representación-aprendizaje con múltiples niveles de representación, obtenidos componiendo módulos simples pero no lineales que transforman la representación a un nivel (iniciando con la entrada cruda) en otra representación a un nivel superior, ligeramente más abstracto.” [1]

—

Dado que el entrenamiento requiere a menudo muchos datos y recursos —y normalmente no se dispone de ellos en cantidades masivas—, las grandes empresas o institutos pueden entrenar modelos en grandes conjuntos de datos.

“El transfer learning es la capacidad de un sistema para reconocer y aplicar conocimientos y habilidades adquiridos en tareas previas a nuevas tareas.” [2]

El Deep Learning, pues, **no solo** extrae y selecciona características, sino que puede implementar cada paso de la *pipeline* de ML de forma **end-to-end**. Dicho aprendizaje se define así:

“El end-to-end learning permite que las redes neuronales transformen entradas crudas (como imágenes) a través de una serie de operaciones, culminando en predicciones finales (por ejemplo, probabilidades de clase). Este proceso de transformación completo se optimiza de forma simultánea mediante backpropagation, ajustando conjuntamente los parámetros de todas las capas en función de la pérdida calculada en la salida.” [3]

GENERAR DATOS CON DEEP LEARNING

Hemos avanzado desde clasificar dígitos hasta redactar historias y crear imágenes. La Tabla muestra aplicaciones y plataformas que generan texto, audio, video e imágenes mediante Deep Learning. El lector puede explorarlas para ver el increíble alcance actual de la generación de contenidos:





Modality	Logo	Name	Functionality	URL
Text-to-text Image-to-text		ChatGPT	It allows the user to engage in human-like conversations and accomplish assorted tasks. It can even answer questions and help you in writing text.	https://chatgpt.com/
Text-to-text Image-to-text		Gemini	It can be used to write something new or to rewrite a given piece of text.	https://gemini.google.com/app
Text-to-text Image-to-text Text-to-image Image-to-image Voice-to-text Voice-to-Image		Microsoft Copilot	It also helps in writing, editing, summarizing, and generating content.	https://copilot.microsoft.com/?form=MA13LV#
Text	—	Bert	It is a language model for natural language processing. It can help machines understand the meaning of text using context.	https://huggingface.co/welcome
Text-to-image		Picsart	It converts text into images.	https://picsart.com/ai-image-generator/

Tabla. Herramientas que usan DL para generar texto, audio, video e imágenes (extracto)






Modality	Logo	Name	Functionality	URL
Text-to-image		Canva	It lets you choose image variations based on a preferred look and composition.	https://www.canva.com/ai-image-generator/
Text-to-image		Adobe	It lets us generate images from text.	https://www.adobe.com/products/firefly/features/text-to-image.html
Text-to-speech		ElevenLabs	Create natural AI voices in any language.	https://elevenlabs.io/
Text-to-speech		PlayHT	Ultra-realistic text-to-speech (TTS) voice. Leading AI voice generator.	https://play.ht/
Text-to-video		Invideo AI	Invideo AI serves as an AI video generator that takes your input script and intelligently crafts it into a video.	https://invideo.io/make/add-text-to-video-online/

Tabla. Herramientas que usan DL para generar texto, audio, video e imágenes (extracto)

CONCLUSIÓN

Se presentó un panorama cronológico del Deep Learning: desde el primer modelo computacional inspirado en la estructura neuronal (McCulloch–Pitts) hasta los modelos generativos actuales. En particular, se revisaron neuronas, Redes Neuronales, CNN, modelos de secuencia y herramientas de última generación capaces de hacer de todo, desde escribir textos hasta generar imágenes.

Las máquinas se están volviendo **creativas**, y serán aún más creativas en el futuro.

REFERENCIAS

1. Definición informal de ML.
2. Definición de IA.
3. Definición formal de ML (Tom Mitchell).
4. Minsky, M. & Papert, S. (1969). *Perceptrons*. MIT Press.
5. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors. Nature*, 323(6088), 533–536.

6. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-based learning applied to document recognition. Proceedings of the IEEE*, 86(11), 2278–2324.
7. Hubel, D. H., & Wiesel, T. N. (1962). *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology*, 160, 106–154.
8. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems (NeurIPS)*.
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.