

Reconocimiento Visual con Deep Learning, ¿dónde estamos?

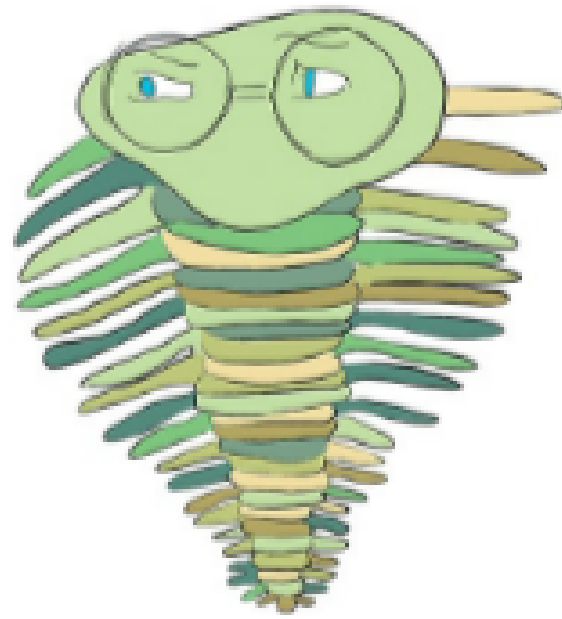


CLASE 1

José M. Saavedra R.
Profesor Asistente

jmsaavedrar@miuandes.cl

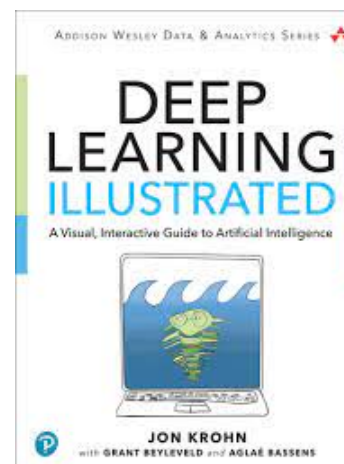
Ed. Ingeniería - Oficina 315



Trilobite*

Visión por Computadora

Rama de la inteligencia artificial (IA) enfocada en el estudio y desarrollo de modelos computacionales capaces de **interpretar nuestro ambiente** a través de información visual (imágenes o videos).



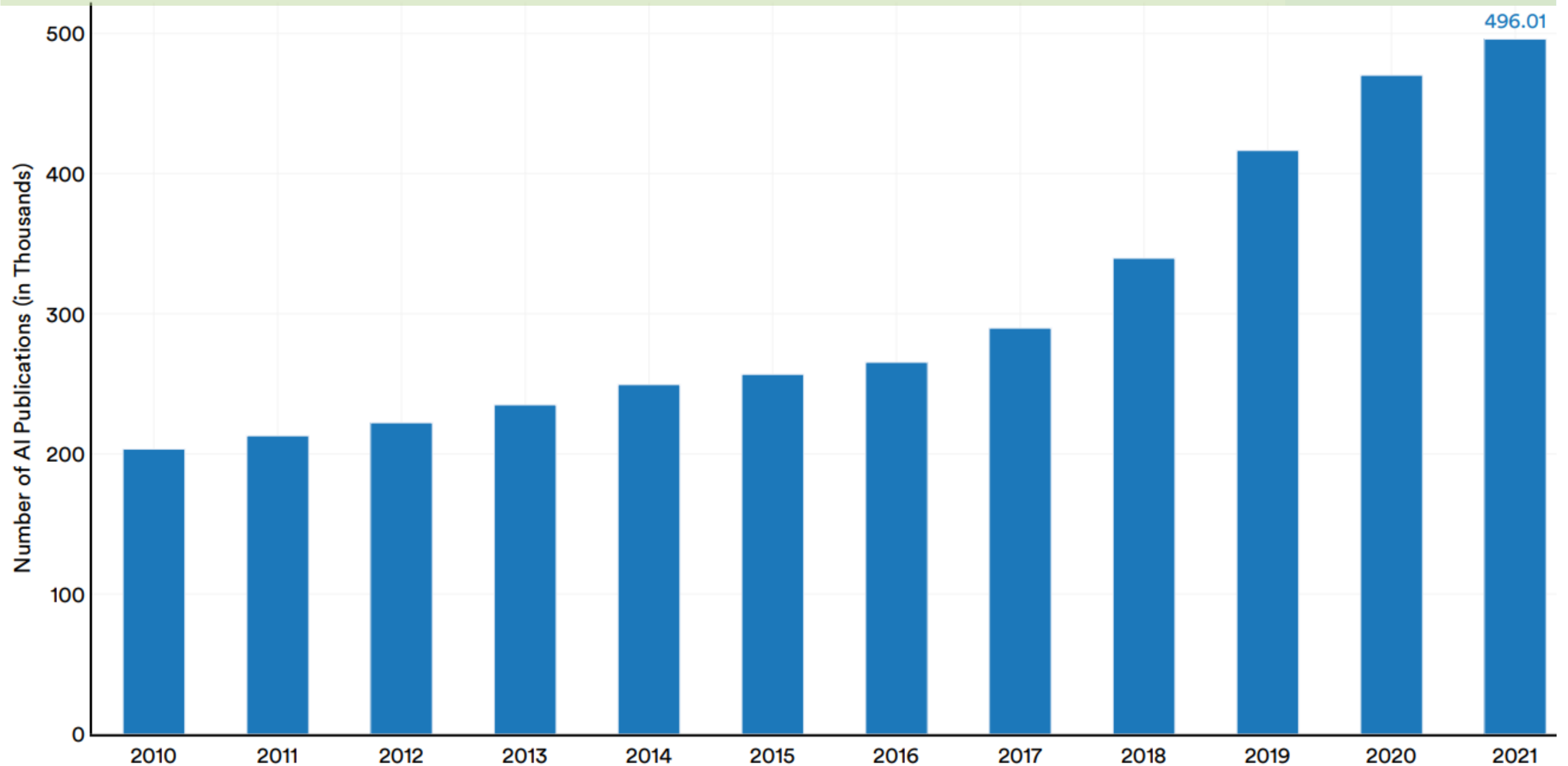
*<https://www.deeplearningillustrated.com/>

Number of AI Publications in the World, 2010–21

Source: Center for Security and Emerging Technology, 2022 | Chart: 2023 AI Index Report



Artificial Intelligence
Index Report 2023



<https://aiindex.stanford.edu/report/>

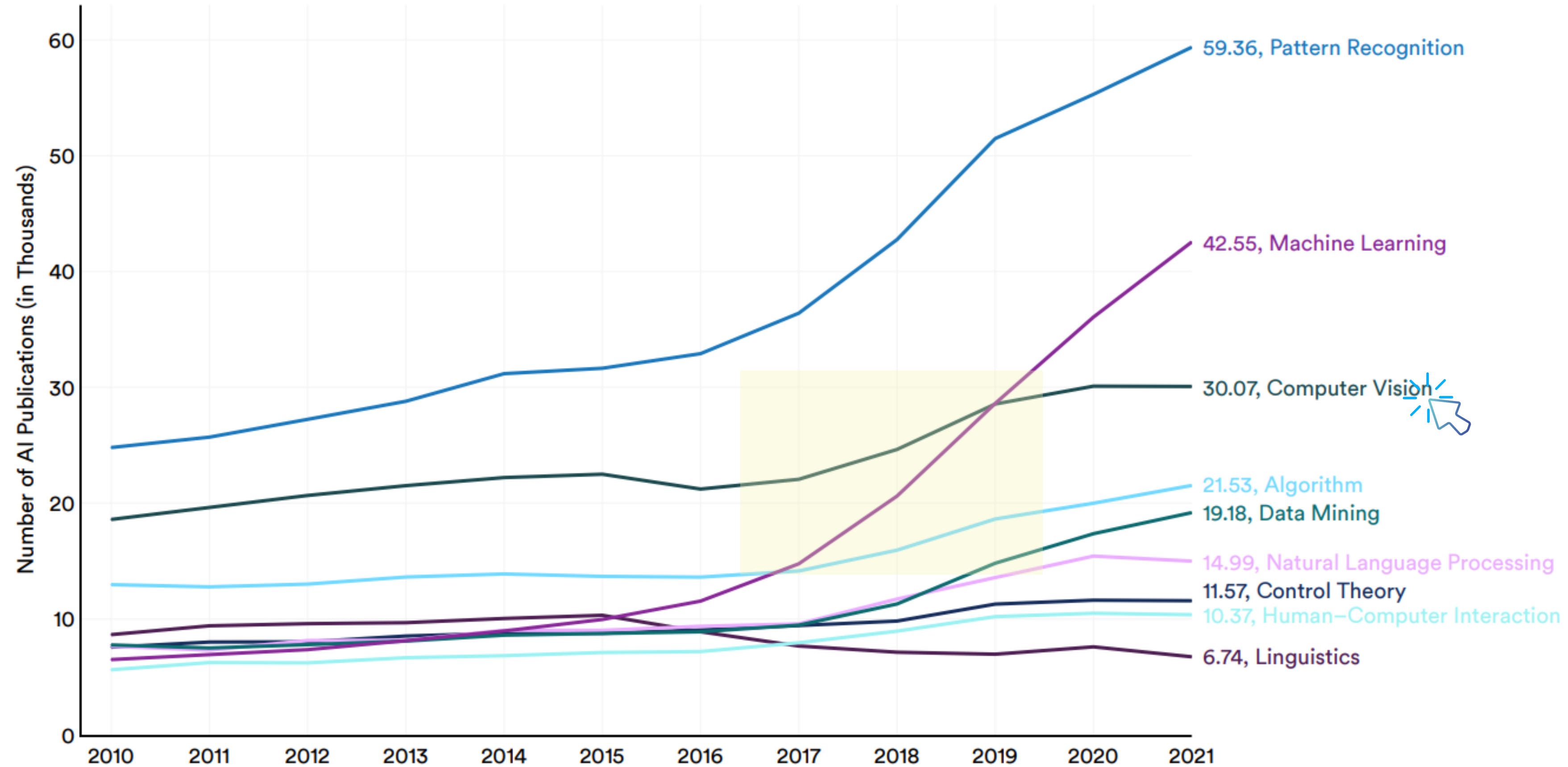
José M. Saavedra

Number of AI Publications by Field of Study (Excluding Other AI), 2010–21

Source: Center for Security and Emerging Technology, 2022 | Chart: 2023 AI Index Report



Artificial Intelligence
Index Report 2023

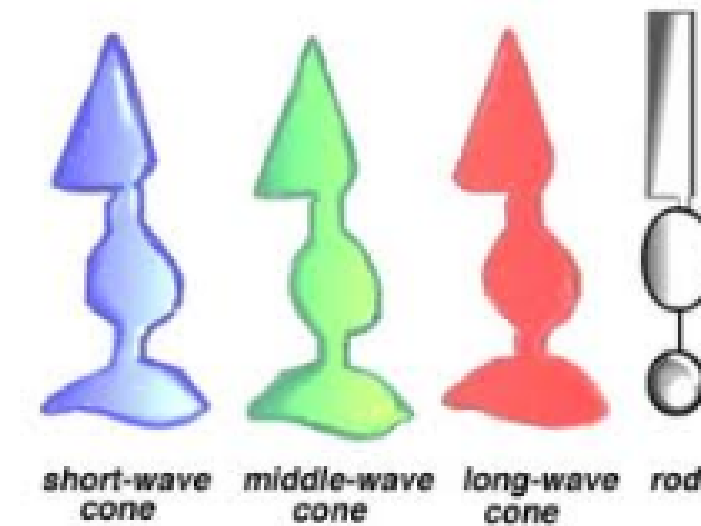


Percepción Visual



- What kind of scene is it?
- What objects appear in the scene?
- Where the objects are located?
- Is it day or night?
- Is it outdoor or indoor?
- Can you provide a brief description?

Percepción Visual

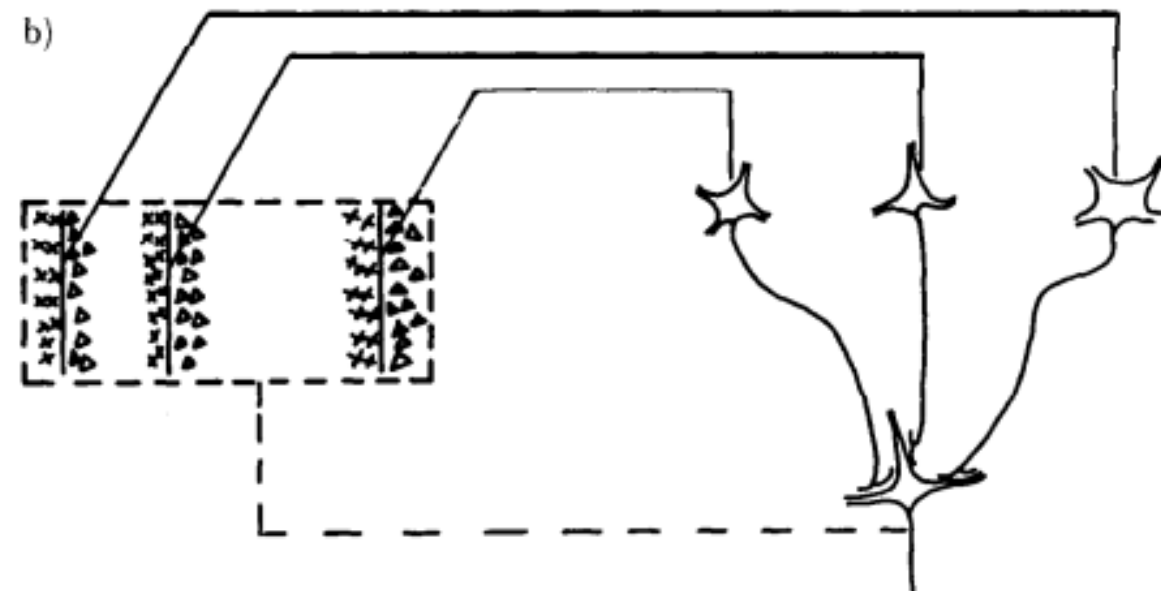
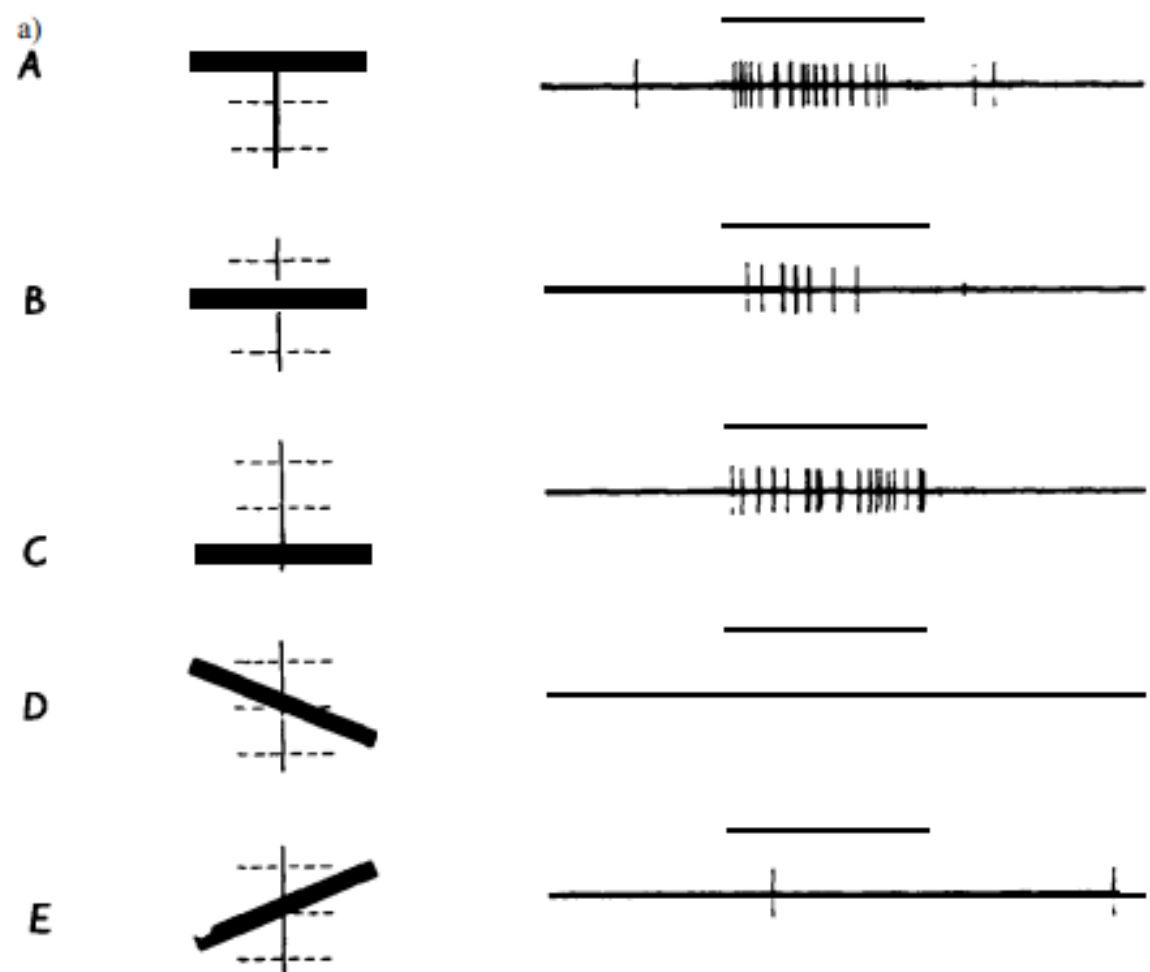


Short ~450nm
Medium ~560nm
Long ~700nm

En la retina encontramos dos tipos de células fotosensibles:

- **Conos:** Visión a color (6 ~ 7 millones).
- **Bastones:** Altamente sensibles a ambientes oscuros (120 millones).

Percepción Visual



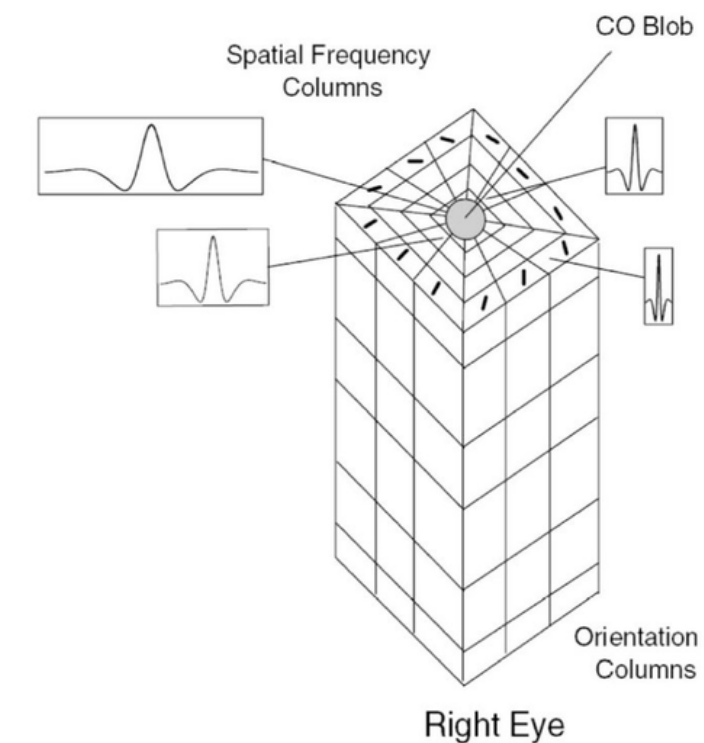
RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX

By D. H. HUBEL AND T. N. WIESEL

*From the Neurophysiology Laboratory, Department of Pharmacology
Harvard Medical School, Boston, Massachusetts, U.S.A.*

(Received 31 July 1961)

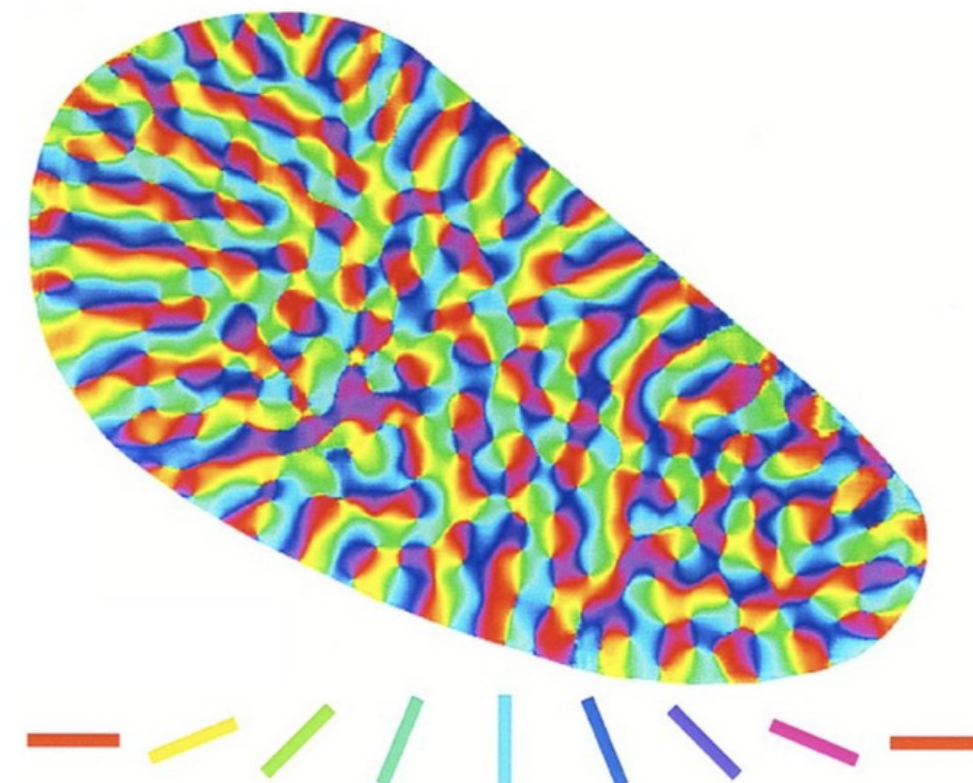
What chiefly distinguishes cerebral cortex from other parts of the central nervous system is the great diversity of its cell types and inter-connexions. It would be astonishing if such a structure did not profoundly modify the response patterns of fibres coming into it. In the cat's visual cortex, the receptive field arrangements of single cells suggest that there is indeed a degree of complexity far exceeding anything yet seen at lower levels in the visual system.



David Hubel



Torsten Wiesel

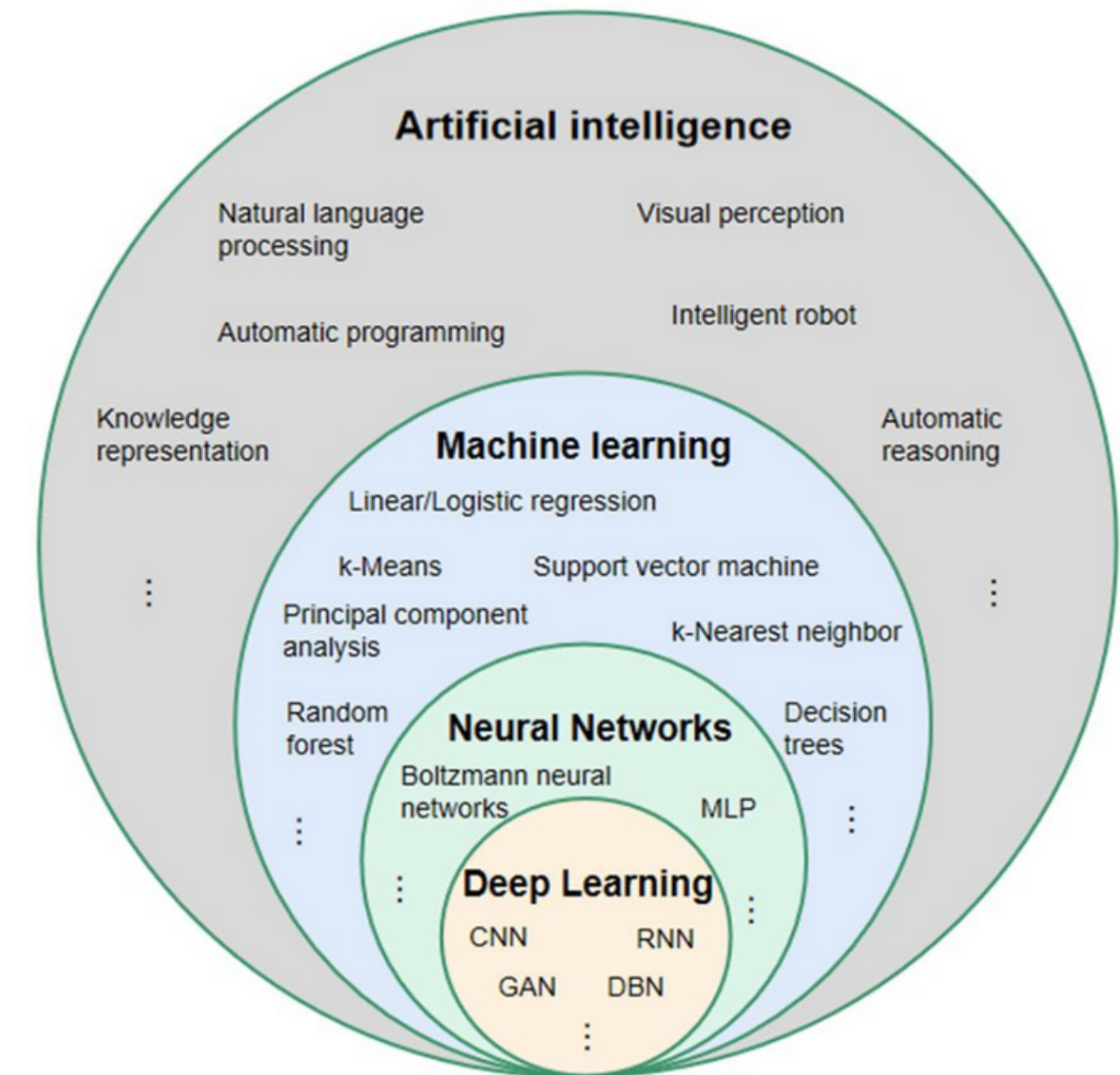


Machine Learning

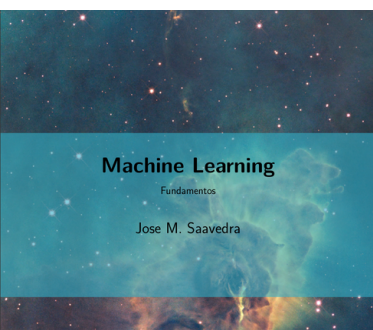
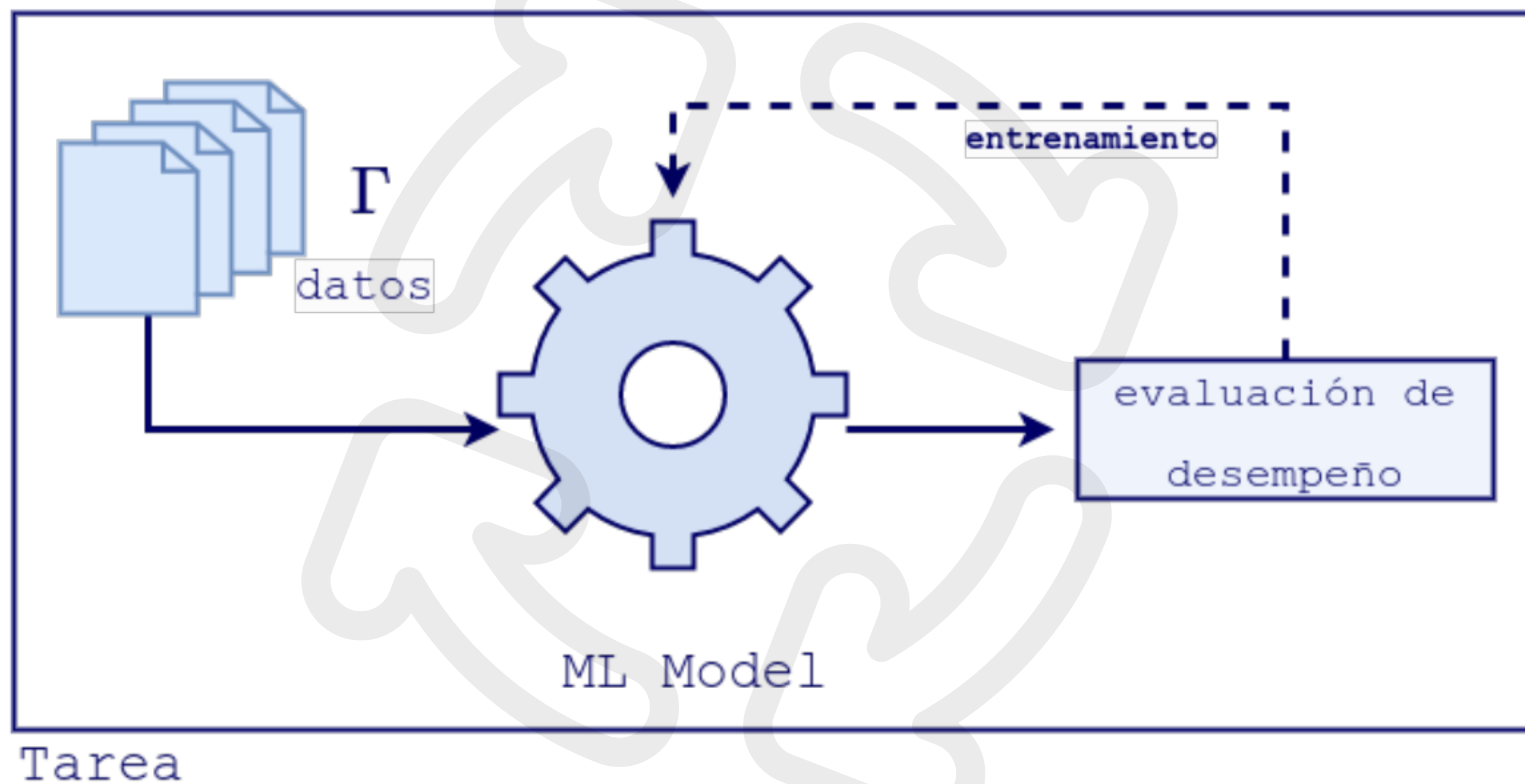
Desarrollo de modelos computacionales capaces de aprender a partir de datos disponibles, para luego realizar predicciones sobre entradas no vistas. Lo que se trata de aprender son patrones, estructuras ocultas o regularidades de los datos, a través de modelos formados por algoritmos y estructuras apropiadas llamados modelos de aprendizaje.

$$y = \mathcal{H}_w(\mathbf{x}) \quad \mathbf{x} \in \mathbb{R}^d$$

← an observation

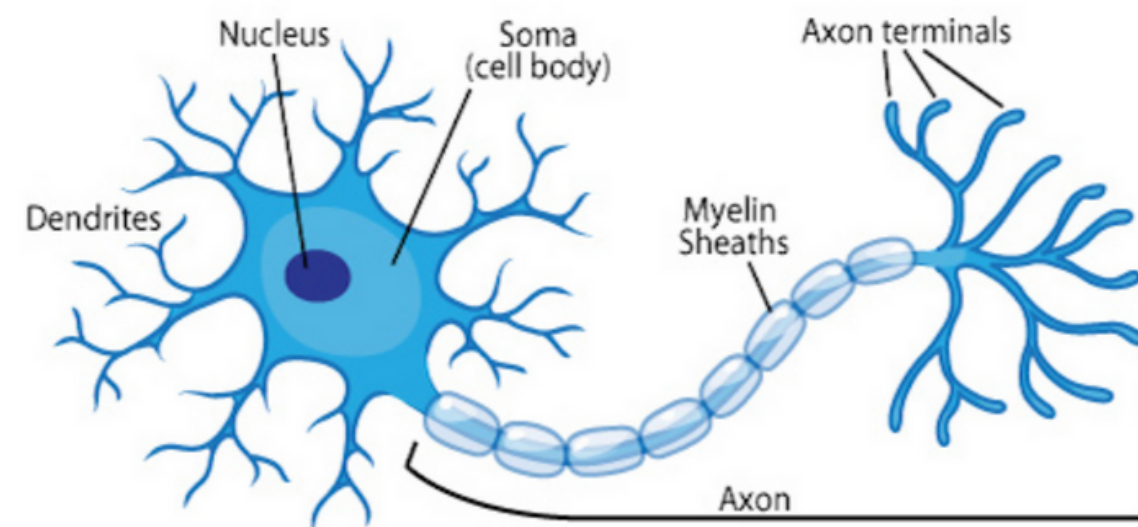


Machine Learning

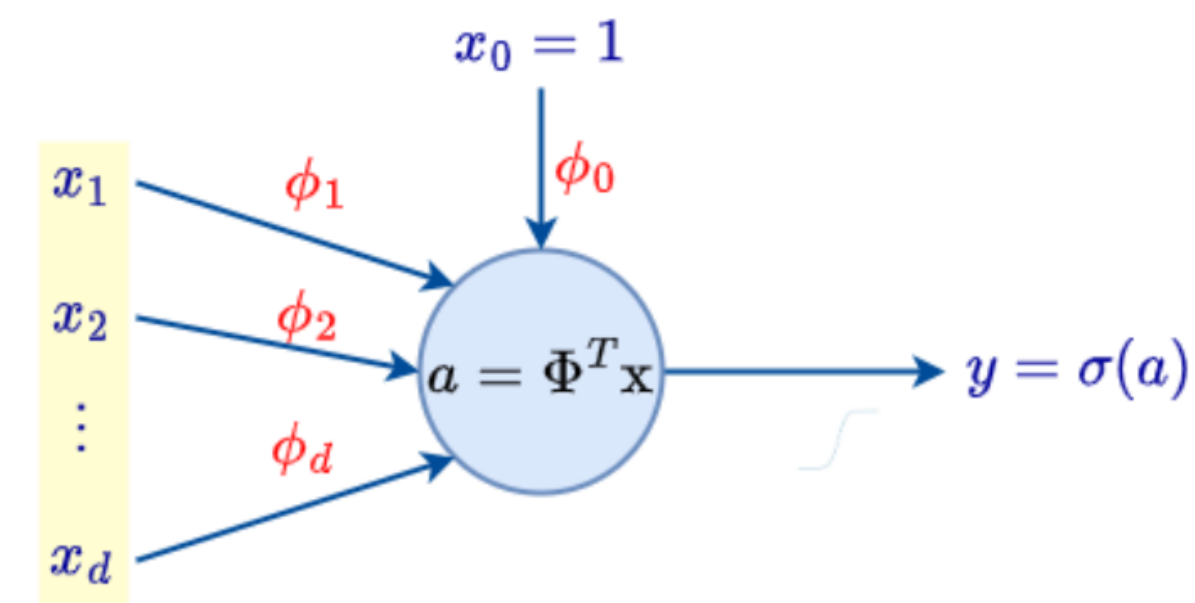


Machine Learning [Redes Neuronales]

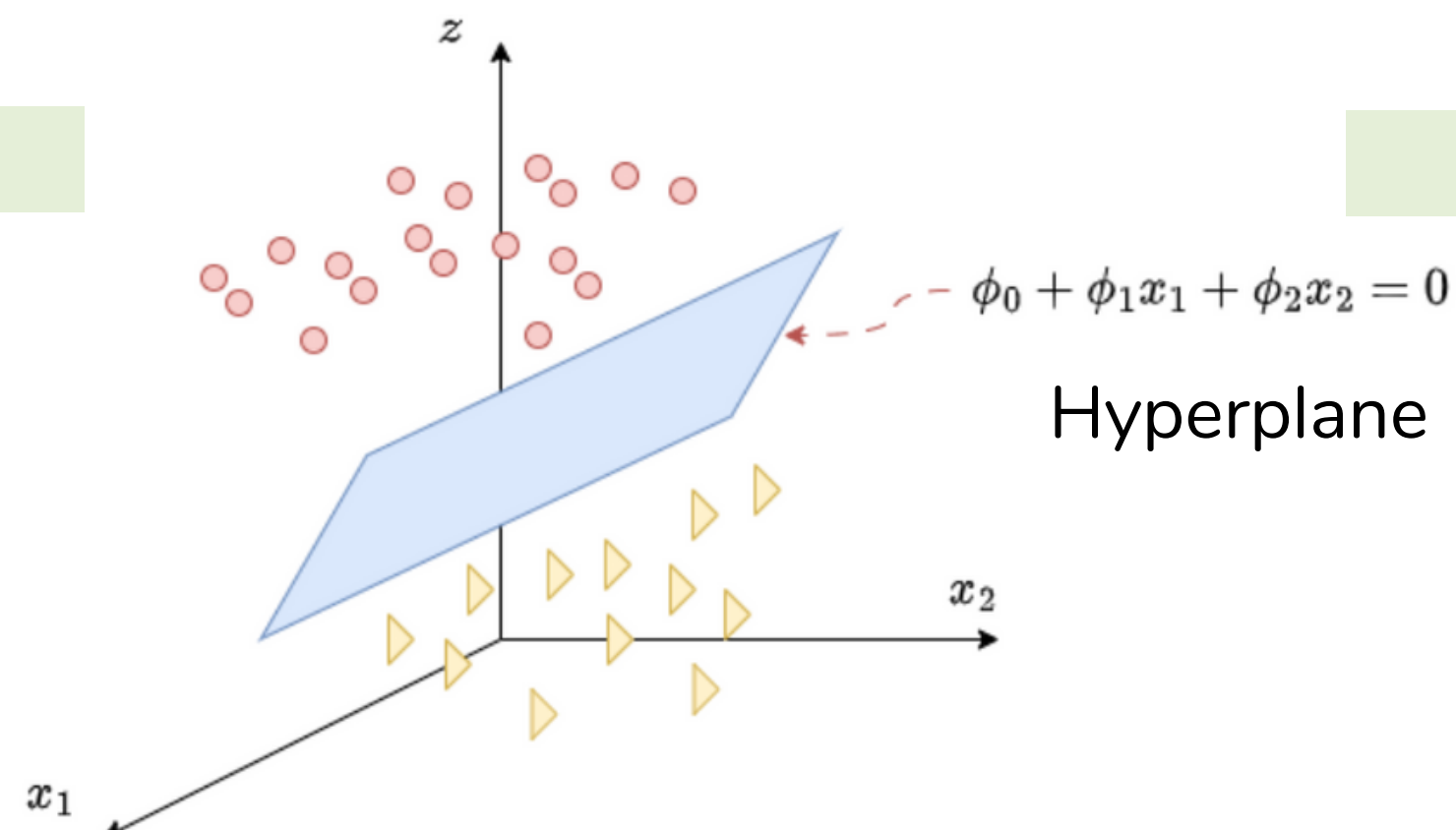
$$\Phi = [\phi_0, \phi_1, \phi_2, \dots, \phi_d] \quad \mathbf{x} = [x_0, x_1, x_2, \dots, x_d]$$



Neuron

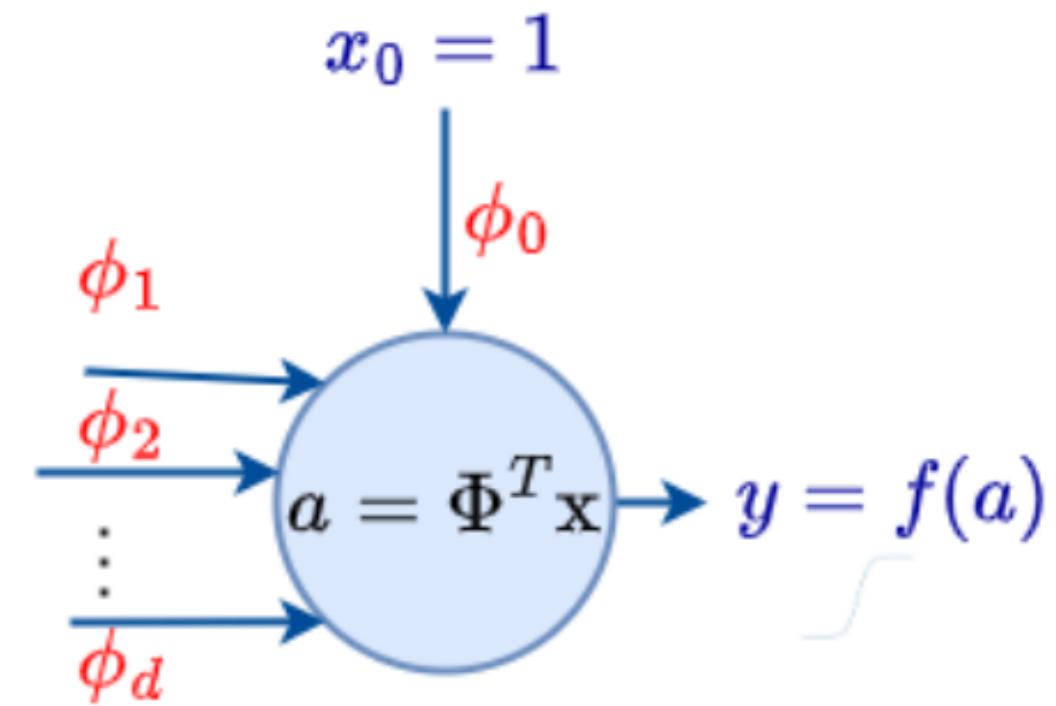
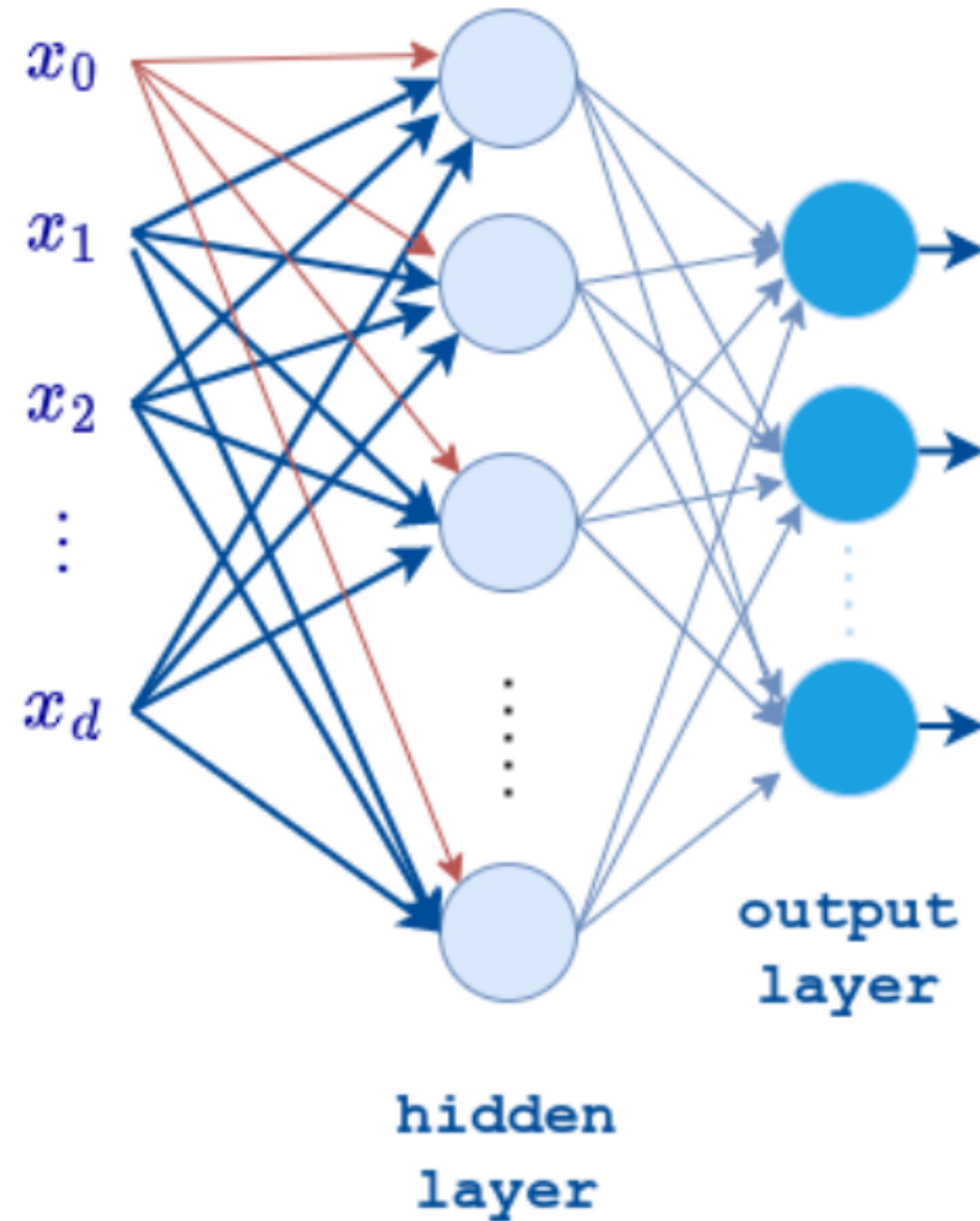


Perceptron



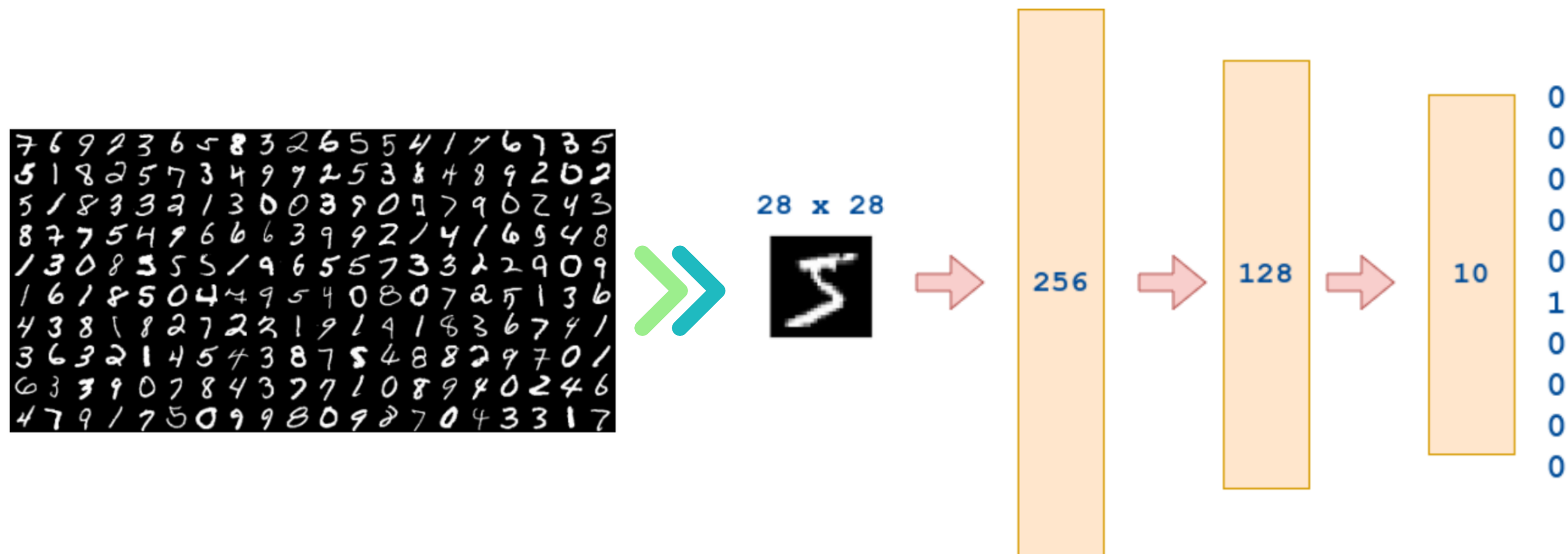
Hyperplane

Machine Learning [Redes Neuronales]



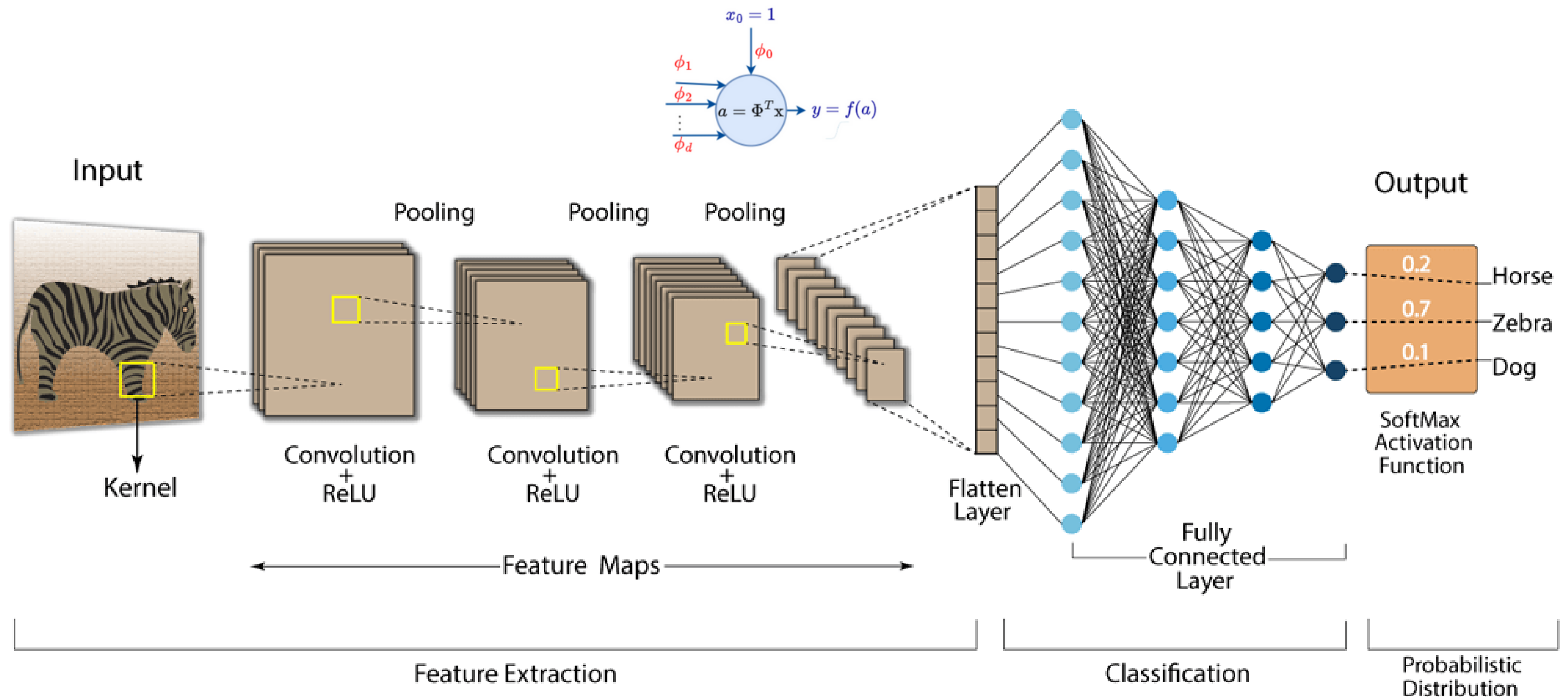
MultiLayer Perceptron [MLP]

Machine Learning [Redes Neuronales]



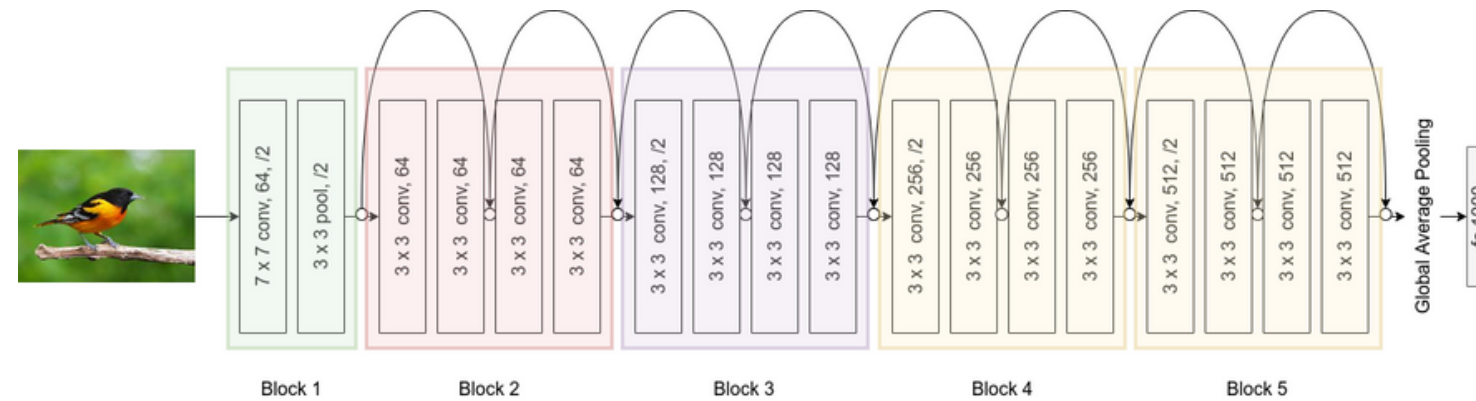
MultiLayer Perceptron [MLP]

Redes Neuronales Convolucionales

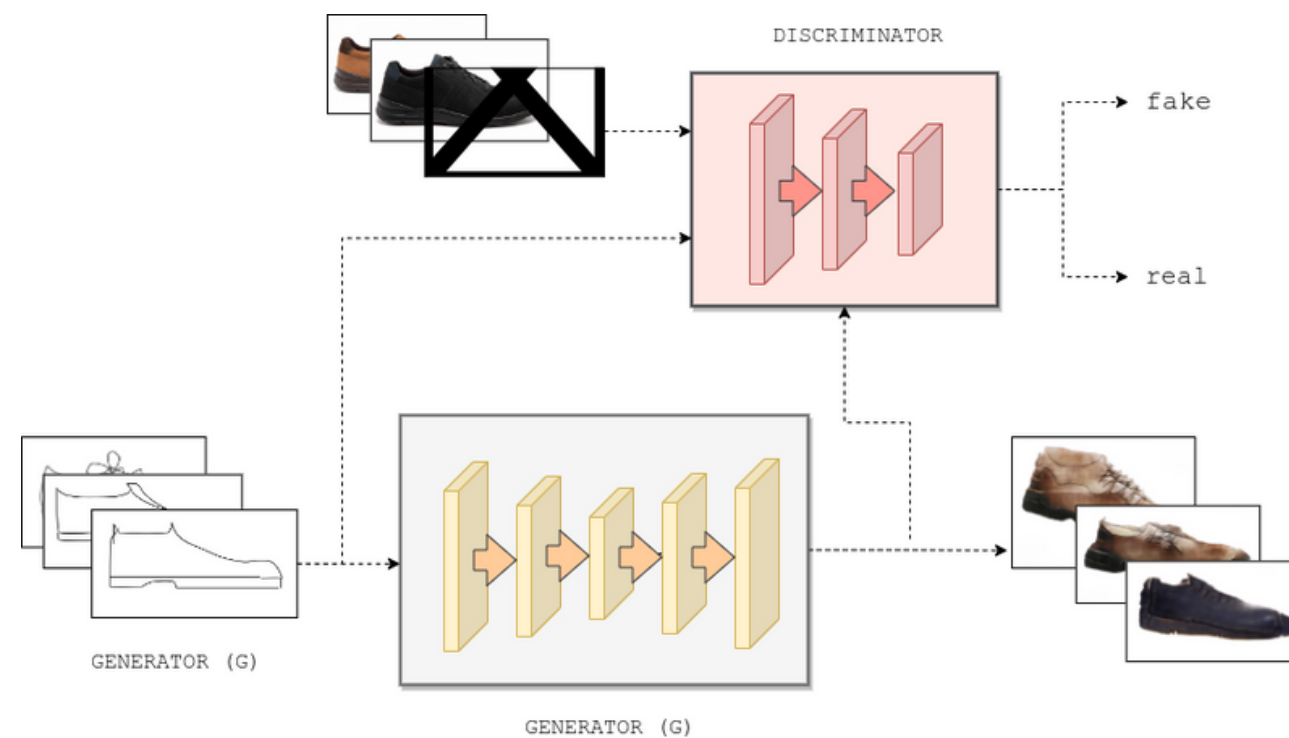


Redes Neuronales Convolucionales

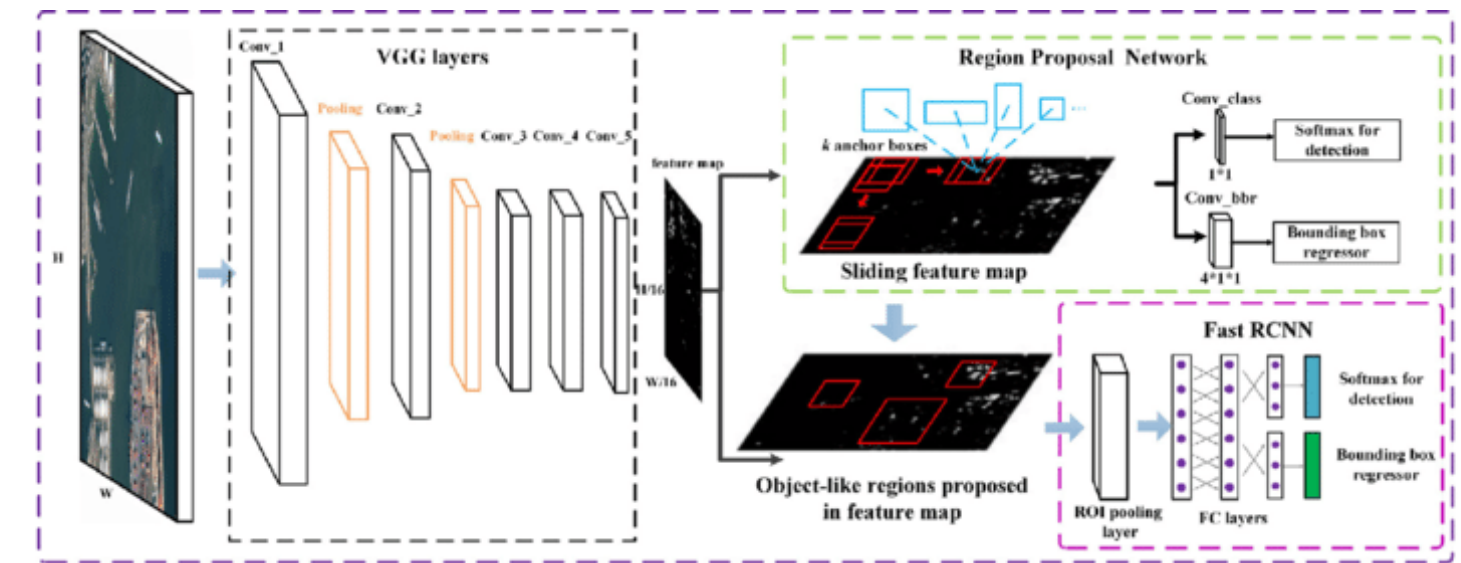
ResNet [backbone]



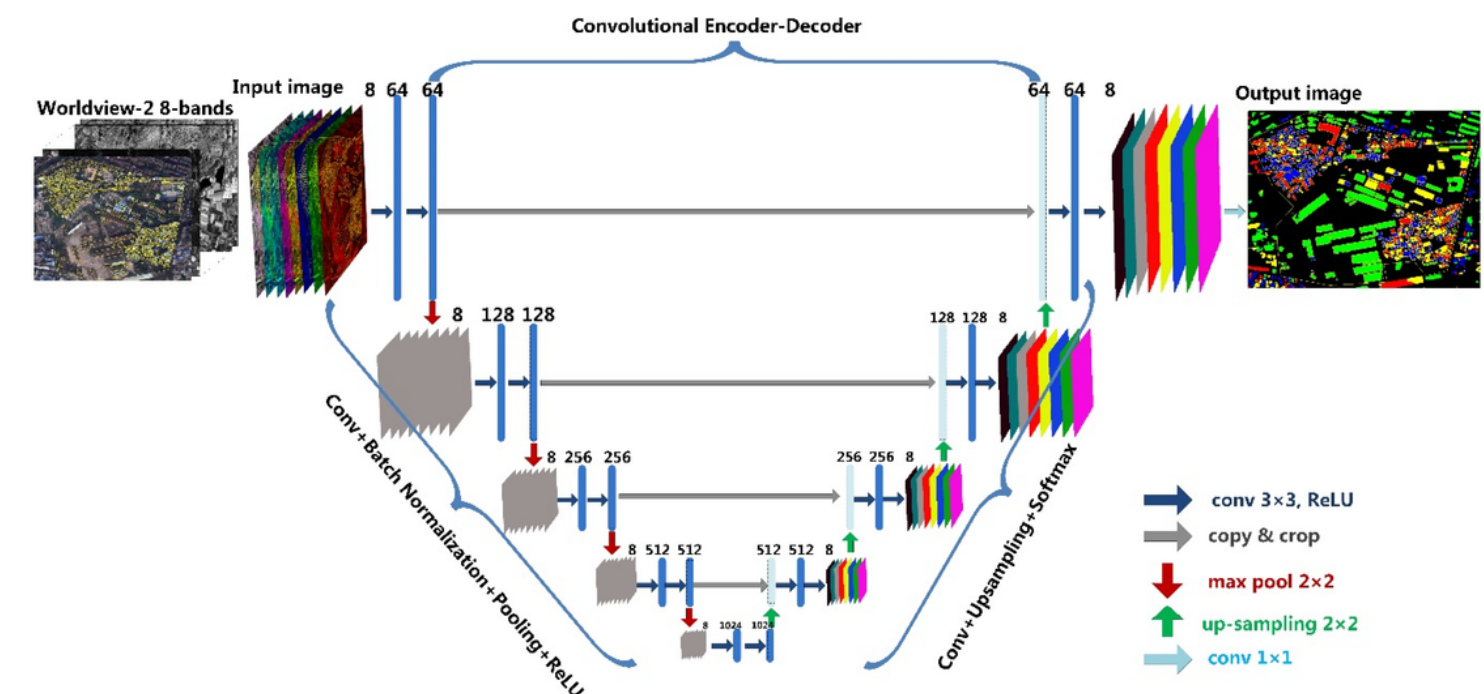
GAN [modelos generativos]



FasterRCNN [detección]



U-Net [segmentación]



2015

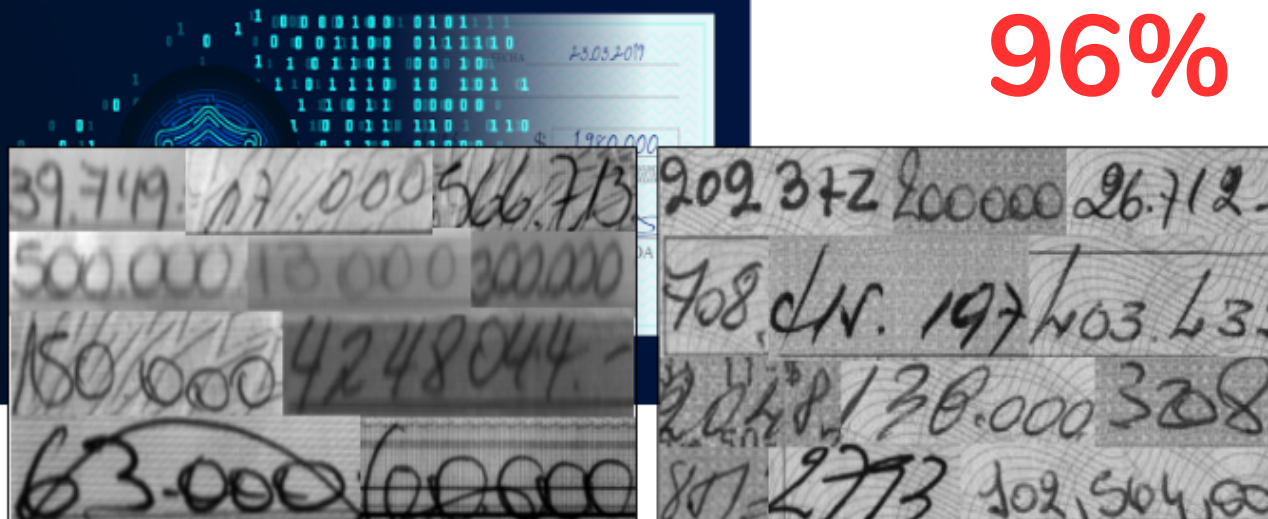
Redes Neuronales Convolucionales



Procesamiento automático y seguro de cheques

Dcōdēr es una solución basada en inteligencia artificial (AI) para procesar y validar documentos bancarios de manera precisa, eficiente y segura.

[Solicitar Demo](#)



96%



impresee CreativeSearch

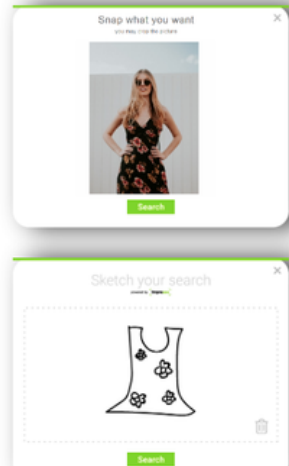
NEW PLUGIN FOR MAGENTO

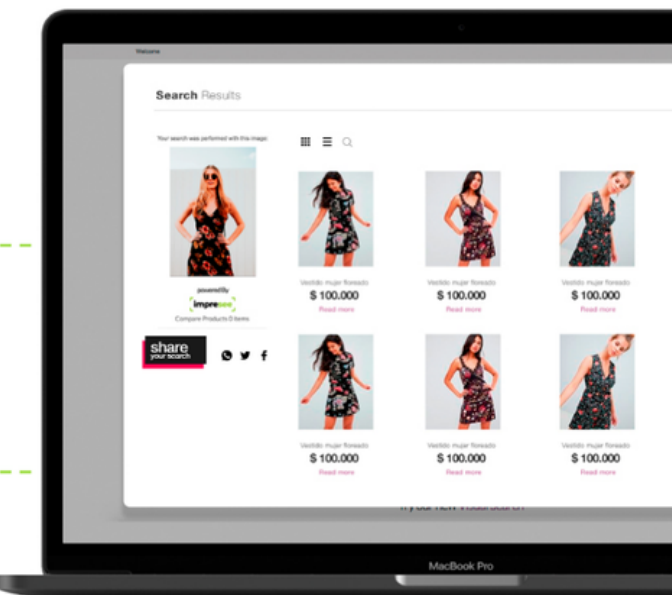


Search By Photo

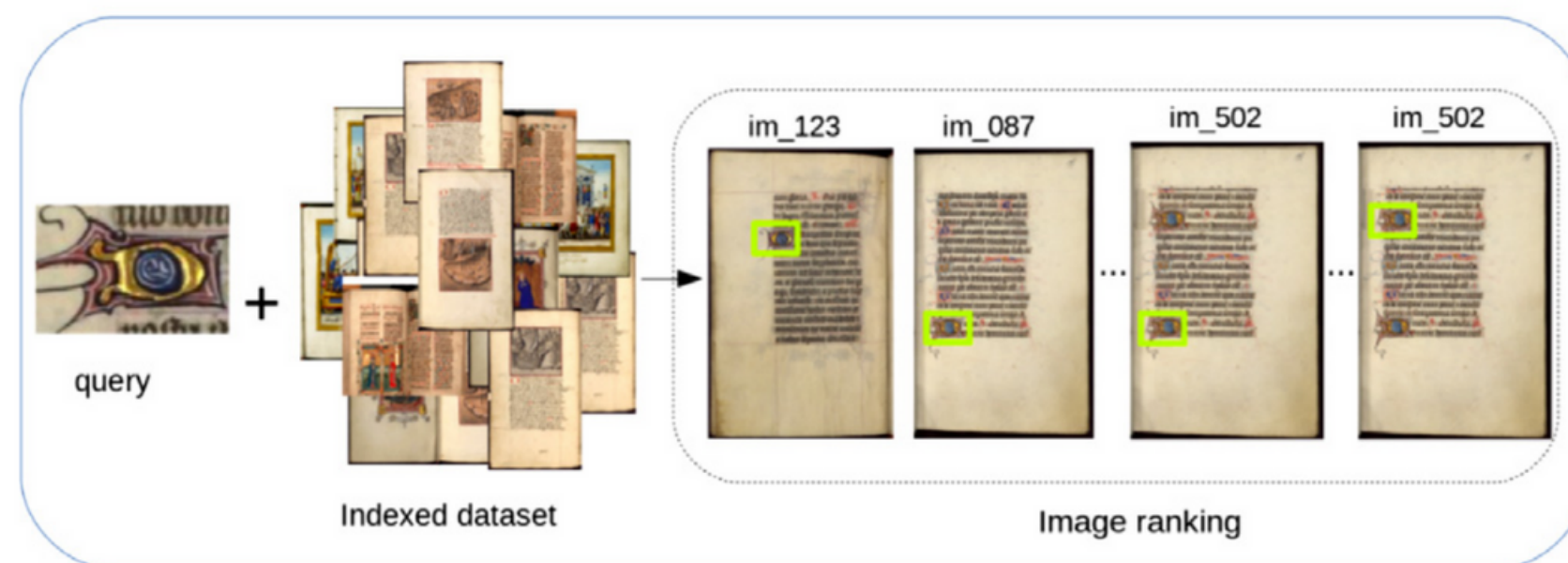


Search By Sketch



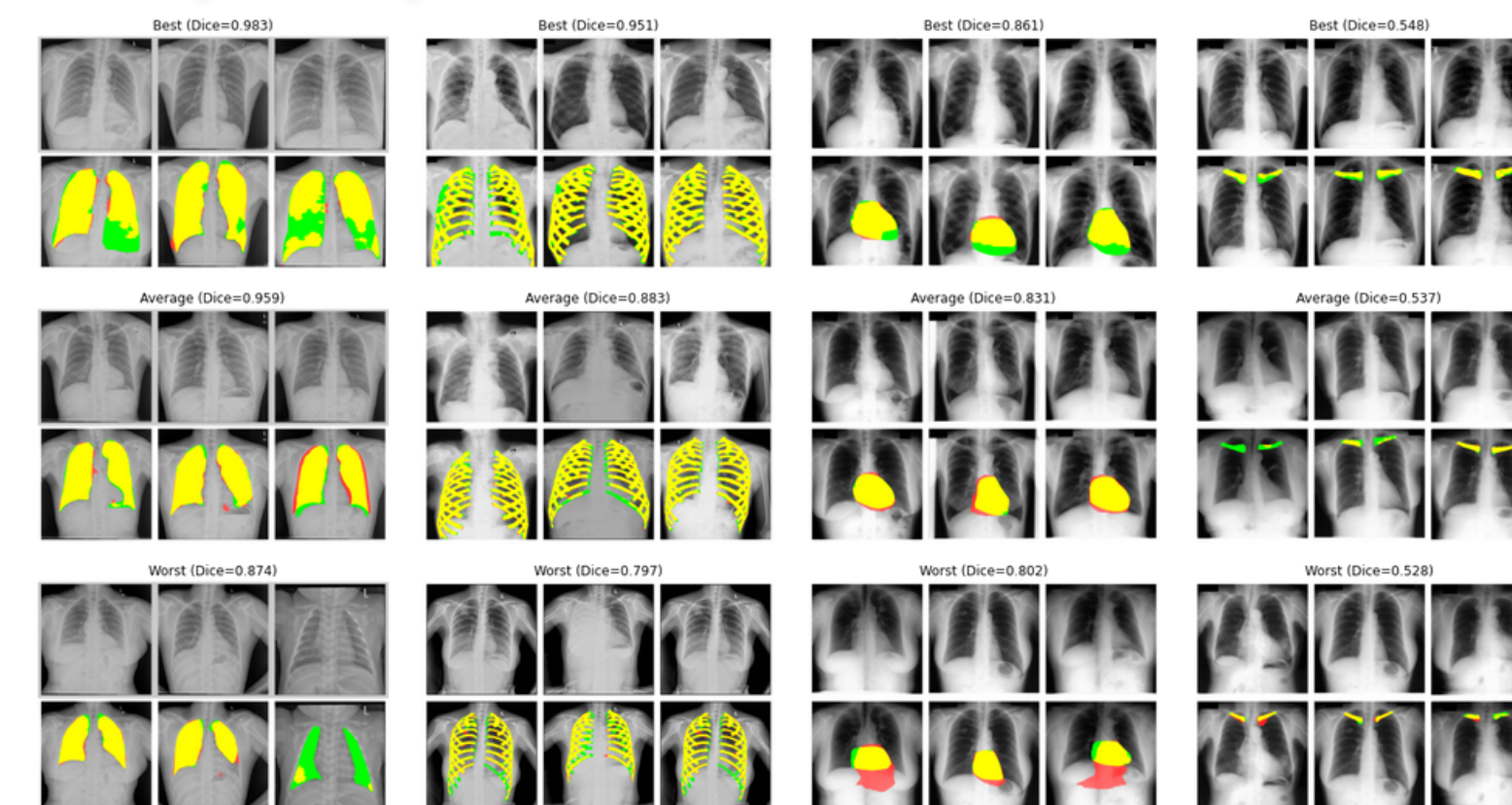


Reconocimiento Montos Manuscritos



Detección de objetos

Búsqueda por Similitud



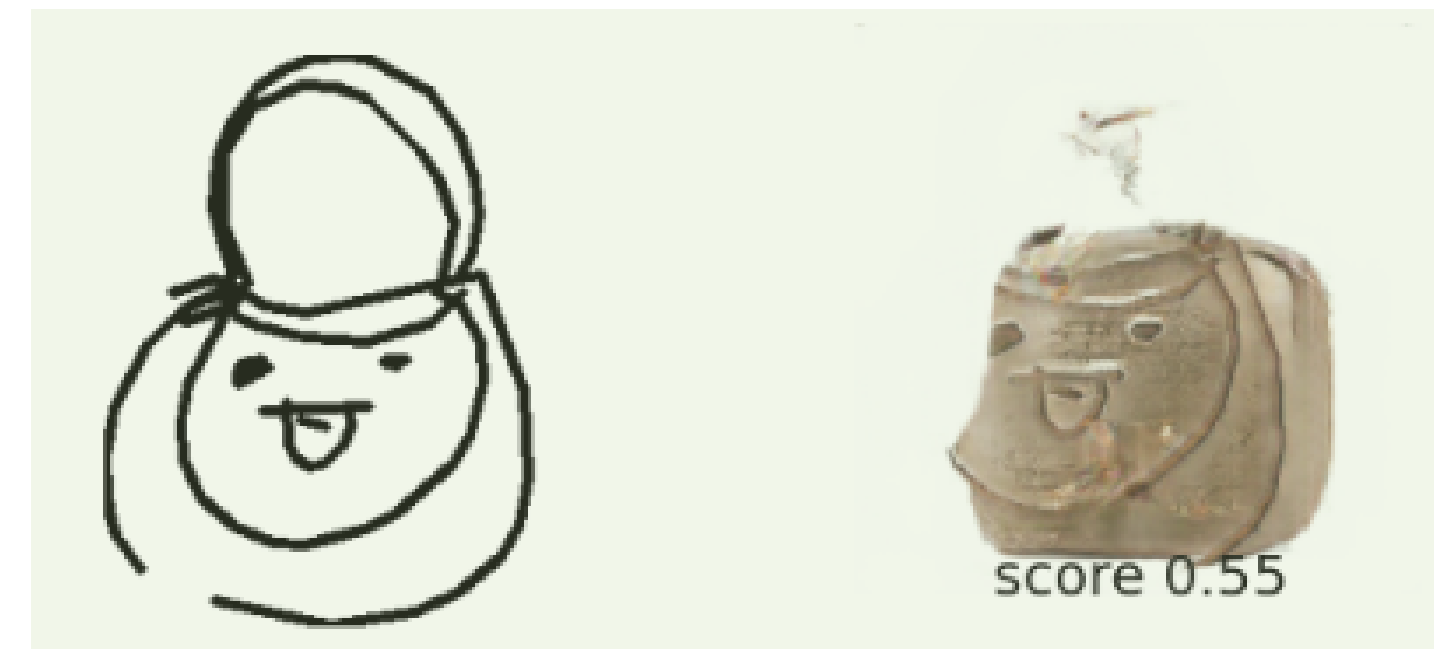
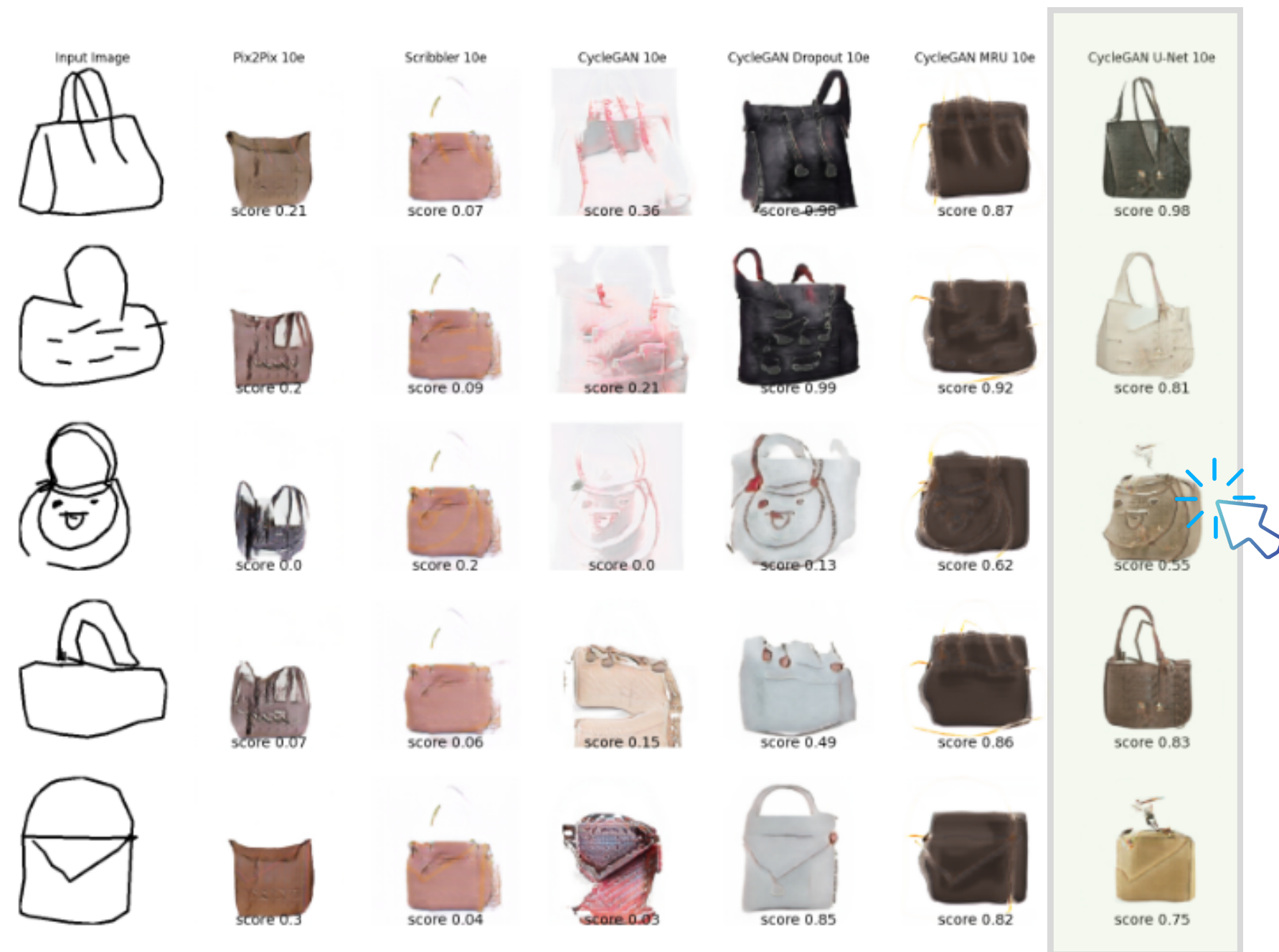
Segmentación

Redes Neuronales Convolucionales

Sketch2Photo Translation

Survey on Sketch-to-Photo Translation

111:19



DATOS

ImageNet 1.3 M

QuickDraw 50 M

SA-1B 11M - 1B

GPT-3 ?TB

OpenImages 15MM boxes, 600 clases

MODELOS

CNN, GAN, DM, Transformers, ViT
self-supervision, multimodall

HARDWARE

GPU, TPU

Aprendizaje de Representaciones (Feature Vector)

Una representación efectiva y eficiente es la clave

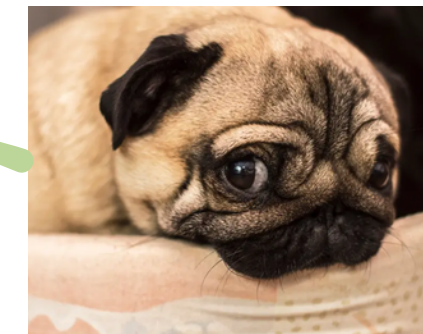
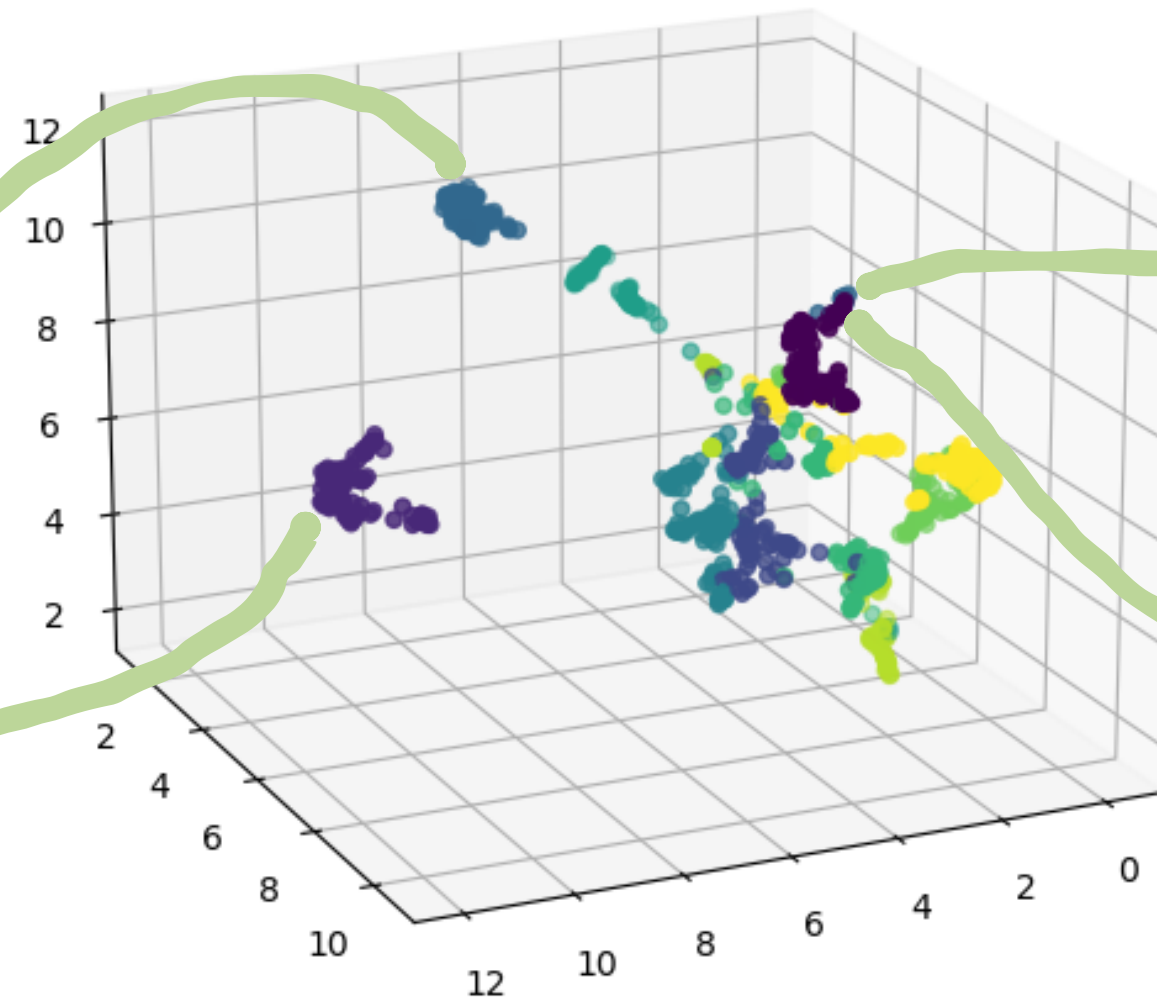
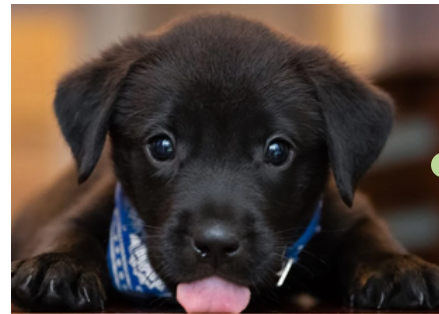
Efectiva > espacio semántico

Eficiente > baja dimensionalidad

feature
vector



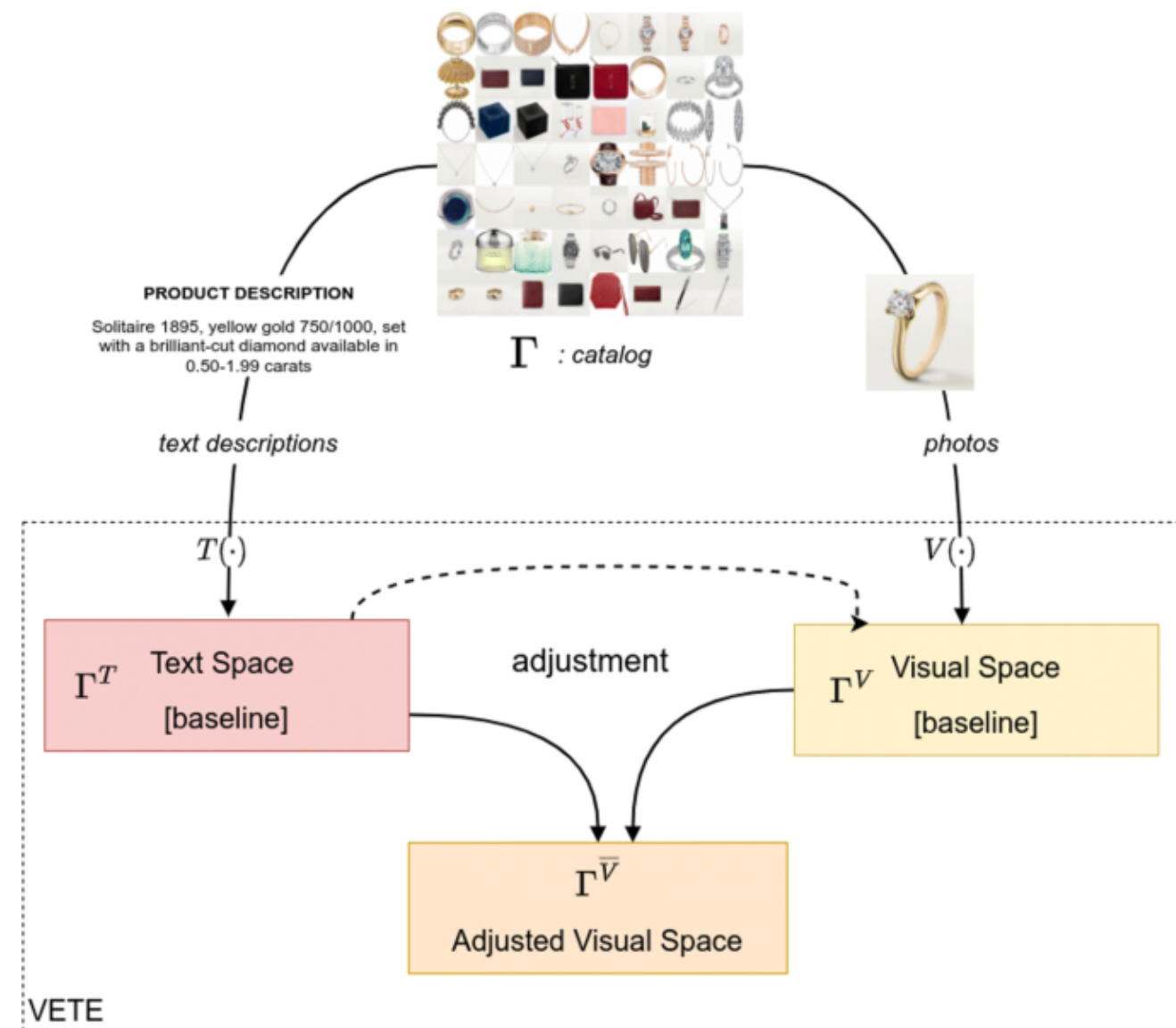
$\mathbf{x} \in \mathbb{R}^d$



¿Cómo?

Multimodalidad

El aprendizaje no es unimodal, aprendemos a partir de diversos y variados estímulos.



[Home](#) > [Multimedia Tools and Applications](#) > [Article](#)

Published: 29 March 2023

VETE: improving visual embeddings through text descriptions for eCommerce search engines

Guillermo Martínez, Jose M. Saavedra & Nils Murrugara-Llerena



(a) Retrieval results in the baseline visual space.

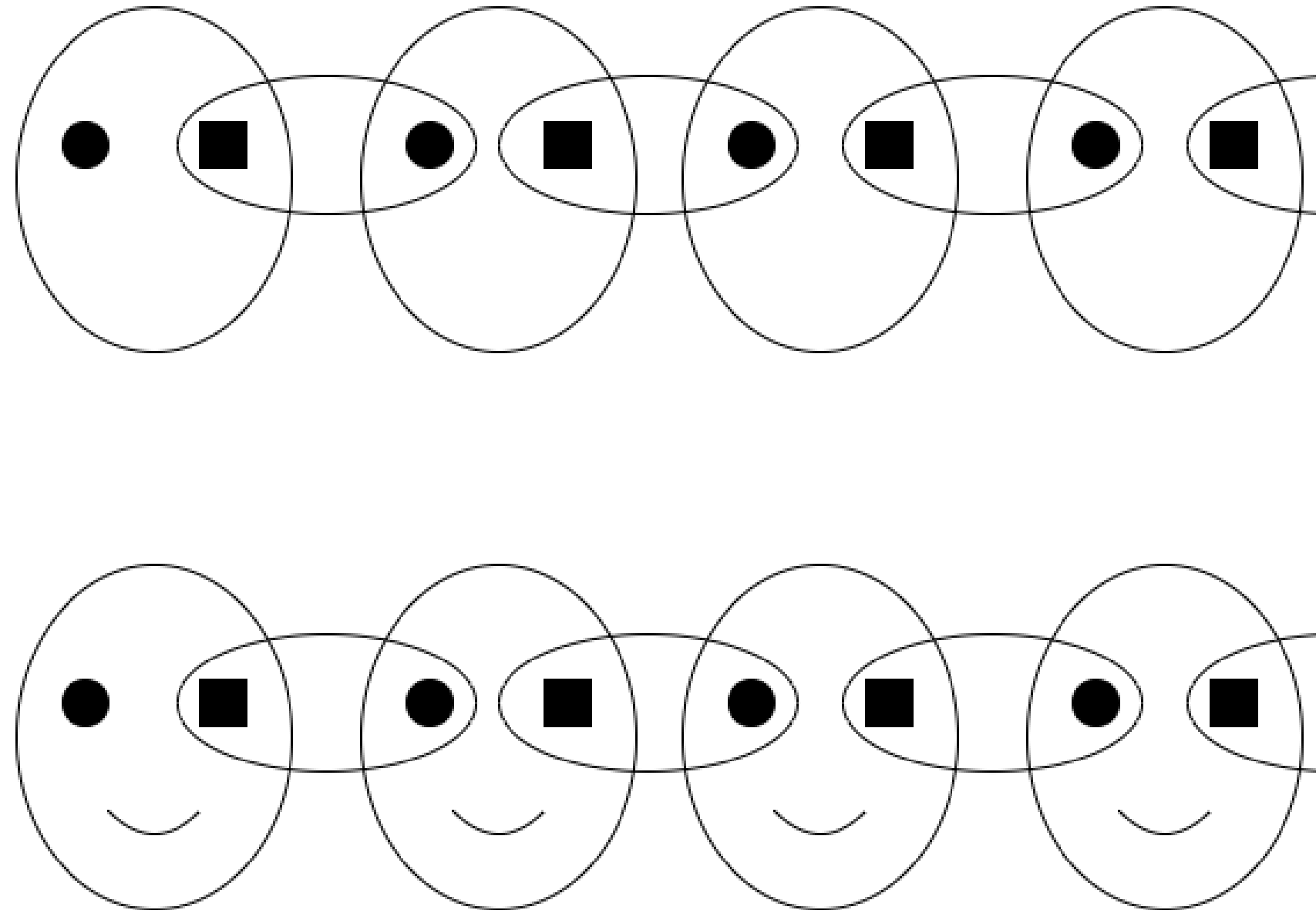


(b) Retrieval results in the visual space adjusted by VETE.

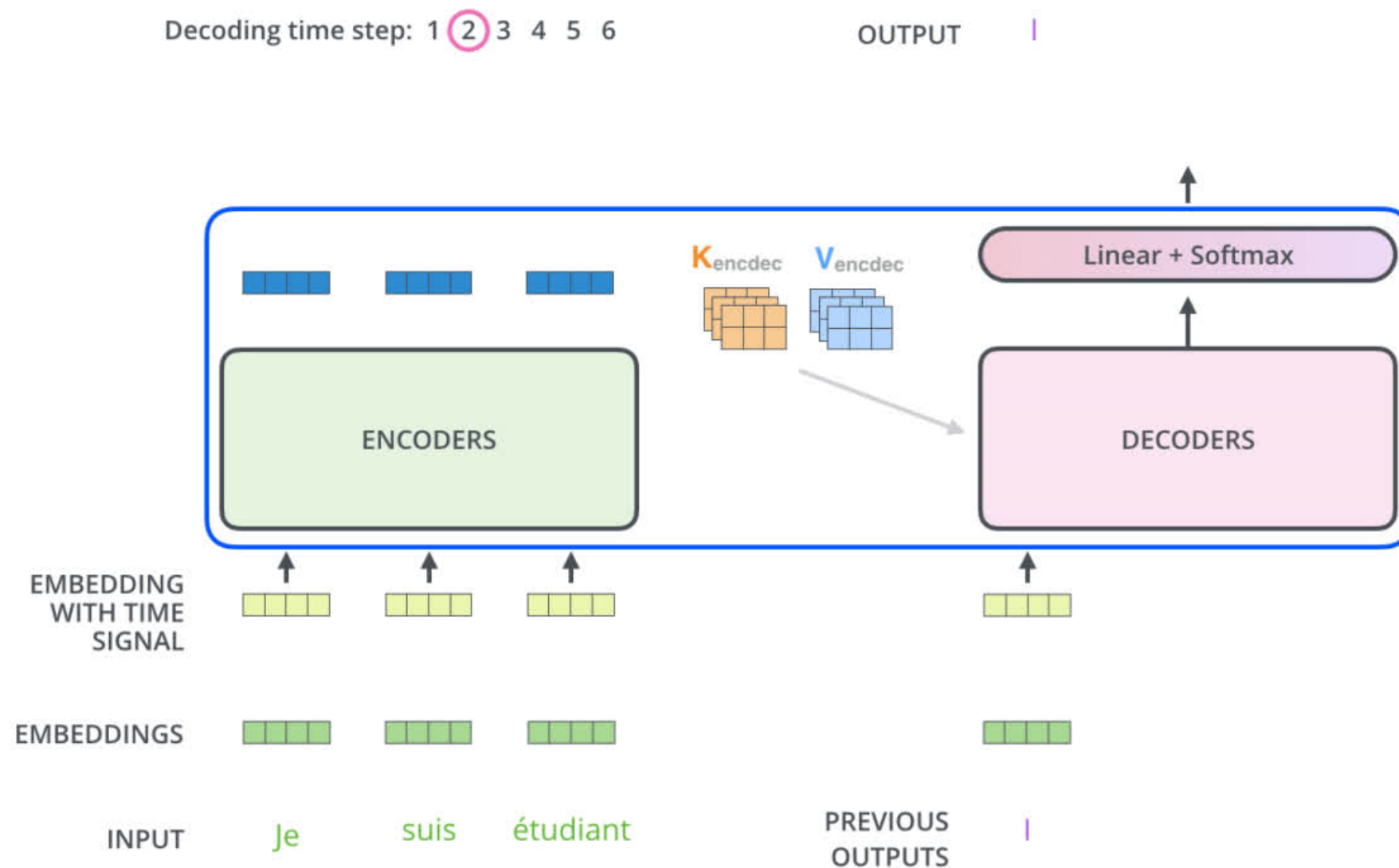
Modelos Atencionales

Perceptual Grouping

percibimos objetos como un todo, más que por sus componentes independientes



Modelos Atencionales



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

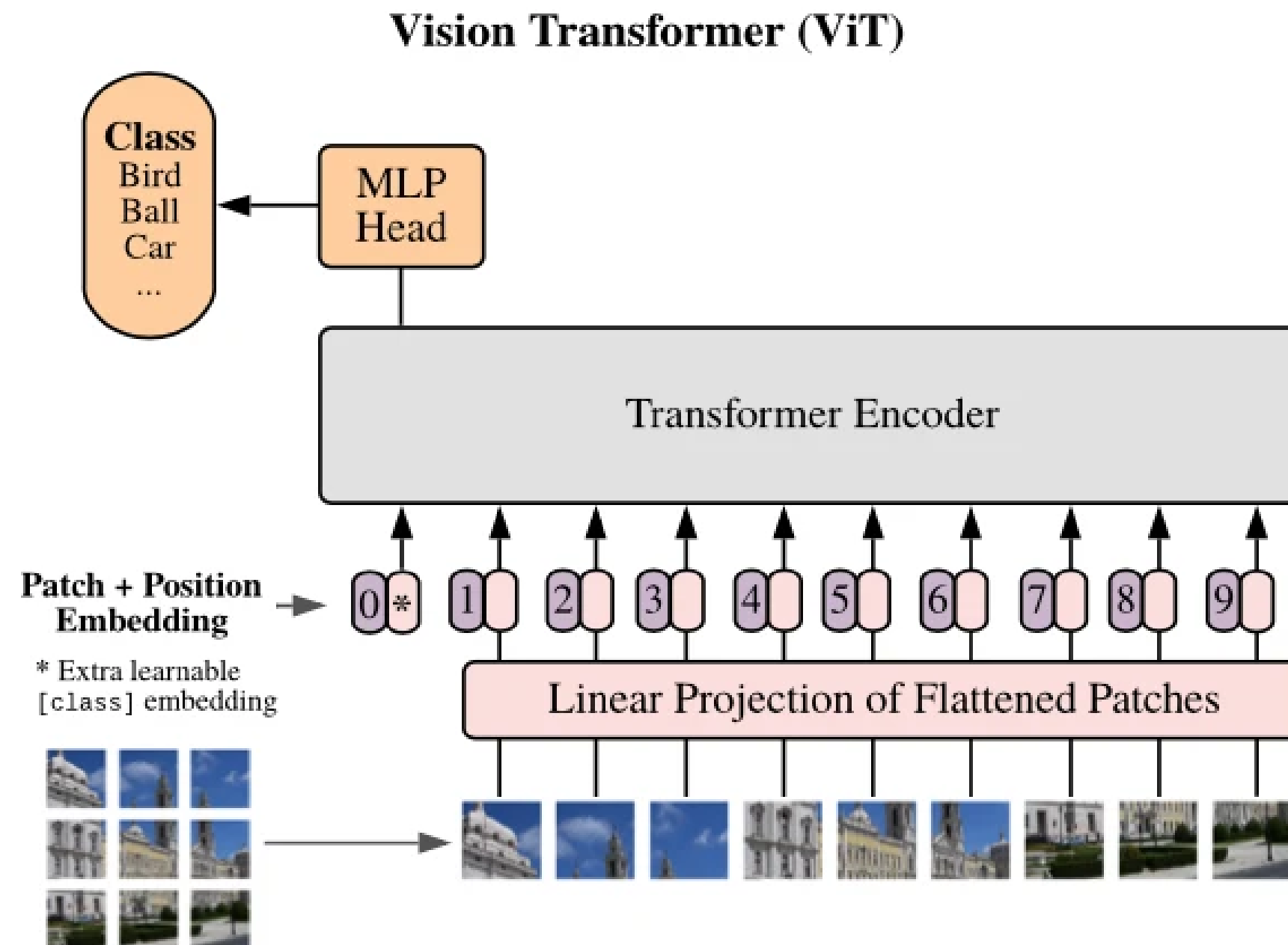
Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

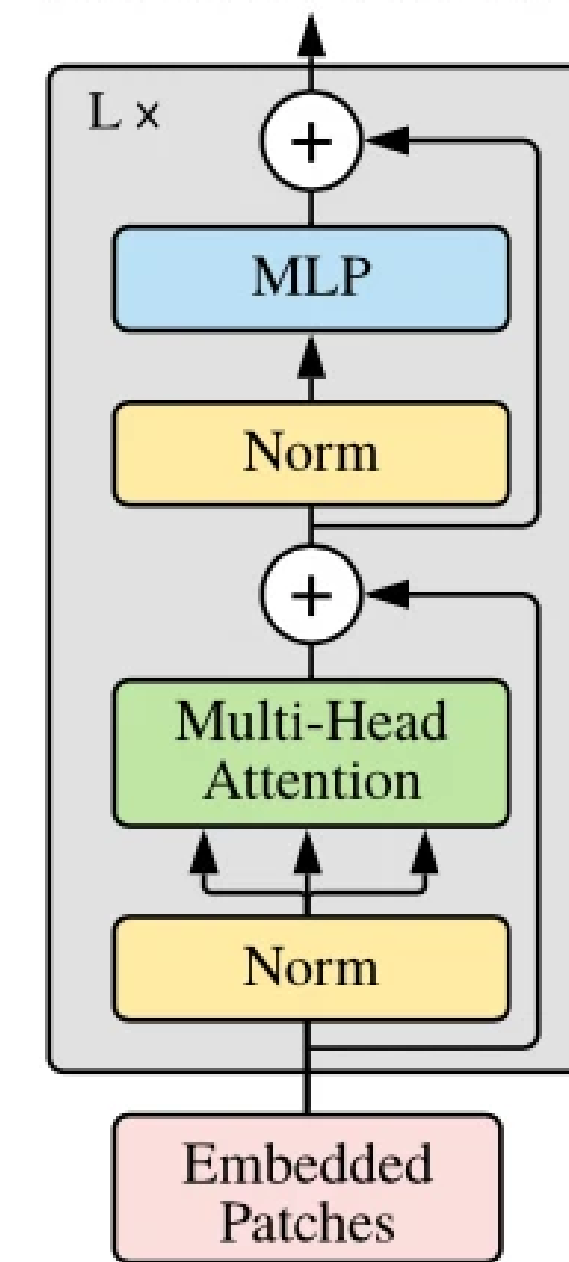
Illia Polosukhin*[‡]
illia.polosukhin@gmail.com

Modelos Atencionales

Vision Transformer ViT

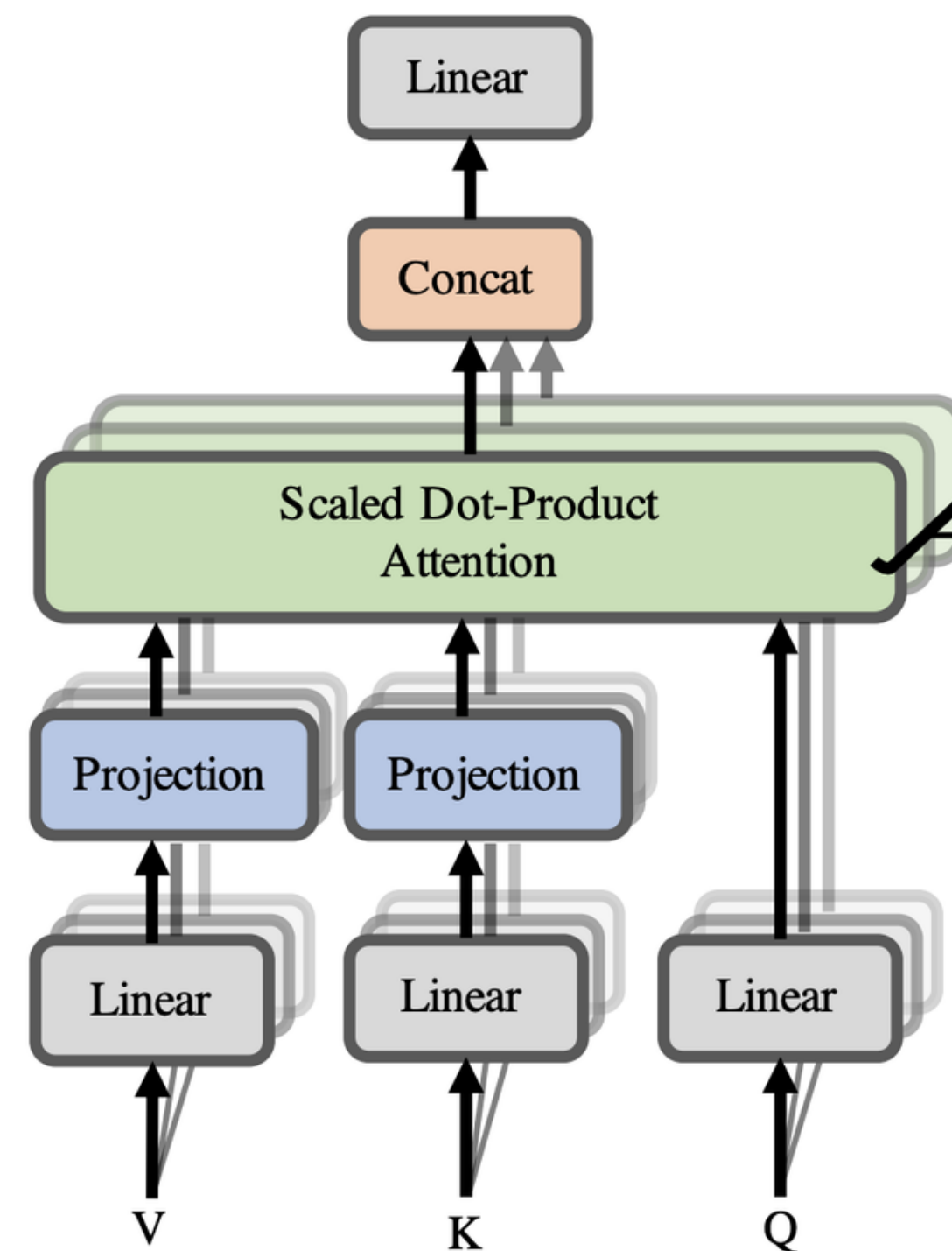
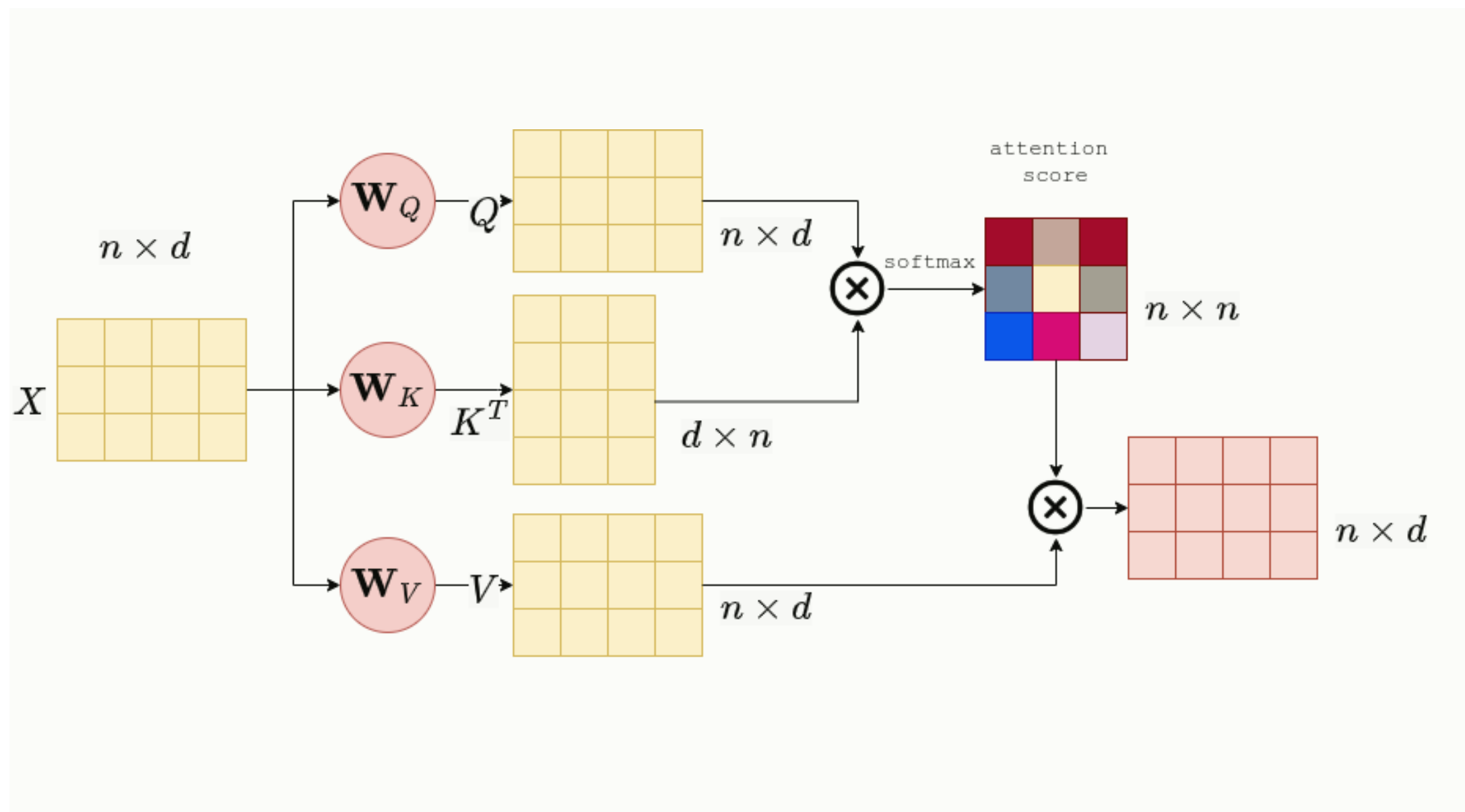


Transformer Encoder



Modelos Atencionales

Vision Transformer ViT



ChatGPT (OpenAI)



tell me about machine learning



Machine learning is a field of computer science and artificial intelligence that focuses on developing algorithms and models that can learn from data, make predictions, and take actions without being explicitly programmed.



The core idea behind machine learning is to train algorithms on a large set of data and use statistical and mathematical techniques to find patterns, make predictions, or identify anomalies in the data. This process involves dividing the data into training, validation, and testing sets, and iteratively adjusting the parameters of the algorithm until it can accurately predict outcomes on new data.

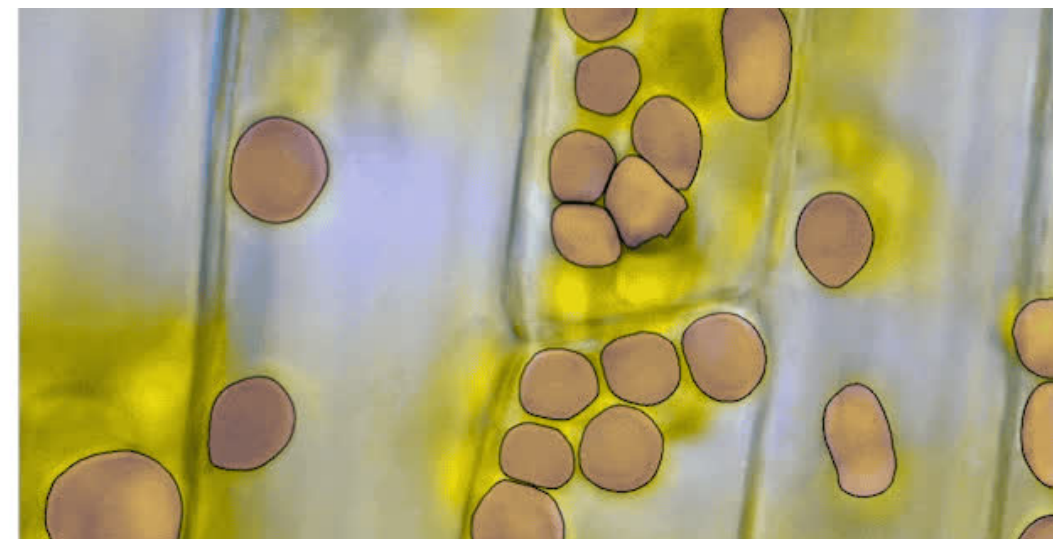
Machine learning algorithms can be broadly classified into three categories: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning involves training an algorithm to predict an output variable based on input variables and a labeled training dataset. The algorithm learns to map inputs to outputs by minimizing the difference between its predictions and the actual values in the training data.

Unsupervised learning involves finding patterns and relationships in unlabeled data, without any prior knowledge of the output variable. Clustering and dimensionality reduction are two common techniques used in unsupervised learning.

Reinforcement learning involves training an agent to make a sequence

SAM (Meta)



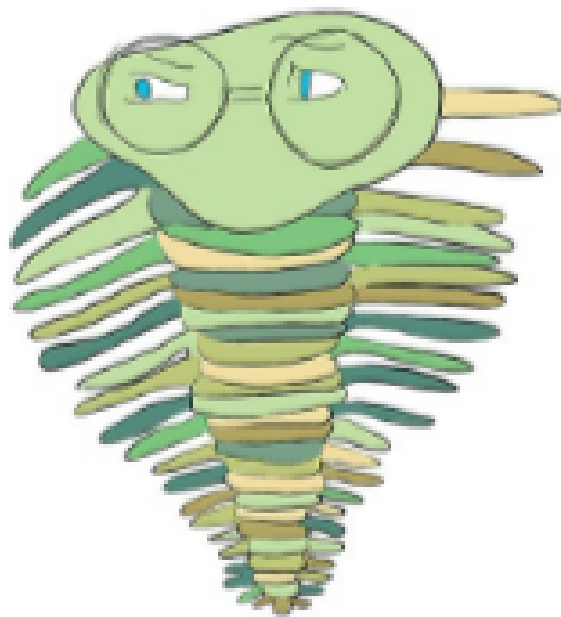
DALL-E (OpenAI)



A banana-shape brown shoe

Desafíos

- Aprovechar modelos genéricos (**Foundation Models**) para mejorar tareas específicas (dominios específicos).
- Aprovechar la gran cantidad de datos multimodales existentes (no necesariamente etiquetados) a través de **estrategias autosupervisadas** para mejorar del desempeño de los modelos ML (de esto hay mucho por hacer en el campo clínico).
- Avanzar en temas de **explicabilidad** (razonamiento, interpretabilidad).
- Incorporar **nuevas modalidades de interacción Human-Computer** (sketch-based understanding).
- Definir regulaciones y **estándares éticos** respecto a integridad, seguridad y diversidad de datos.



Muchas Gracias

jmsaavedrar@miuandes.cl



Clase 2

- Imagen digital y propiedades
- Filtrado lineal y
- Redes convolucionales
- Lab 1: Redes convolucionales