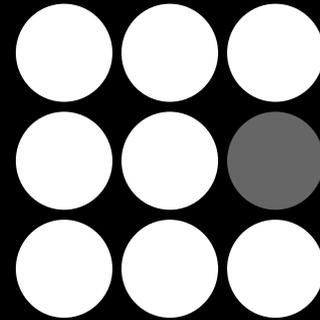


El **criterio**\_ es el  
arte de **\_elegir**



















átomos

bits

los cimientos de los dos universos en los que vivimos

El proceso de digitalización que está  
viviendo la humanidad es un proceso  
de transformación **del átomo al bit**

*Being digital*

Nicholas Negroponte, 1996



Byung-Chul  
Han

# NO-COSAS

QUIEBRAS DEL MUNDO DE HOY



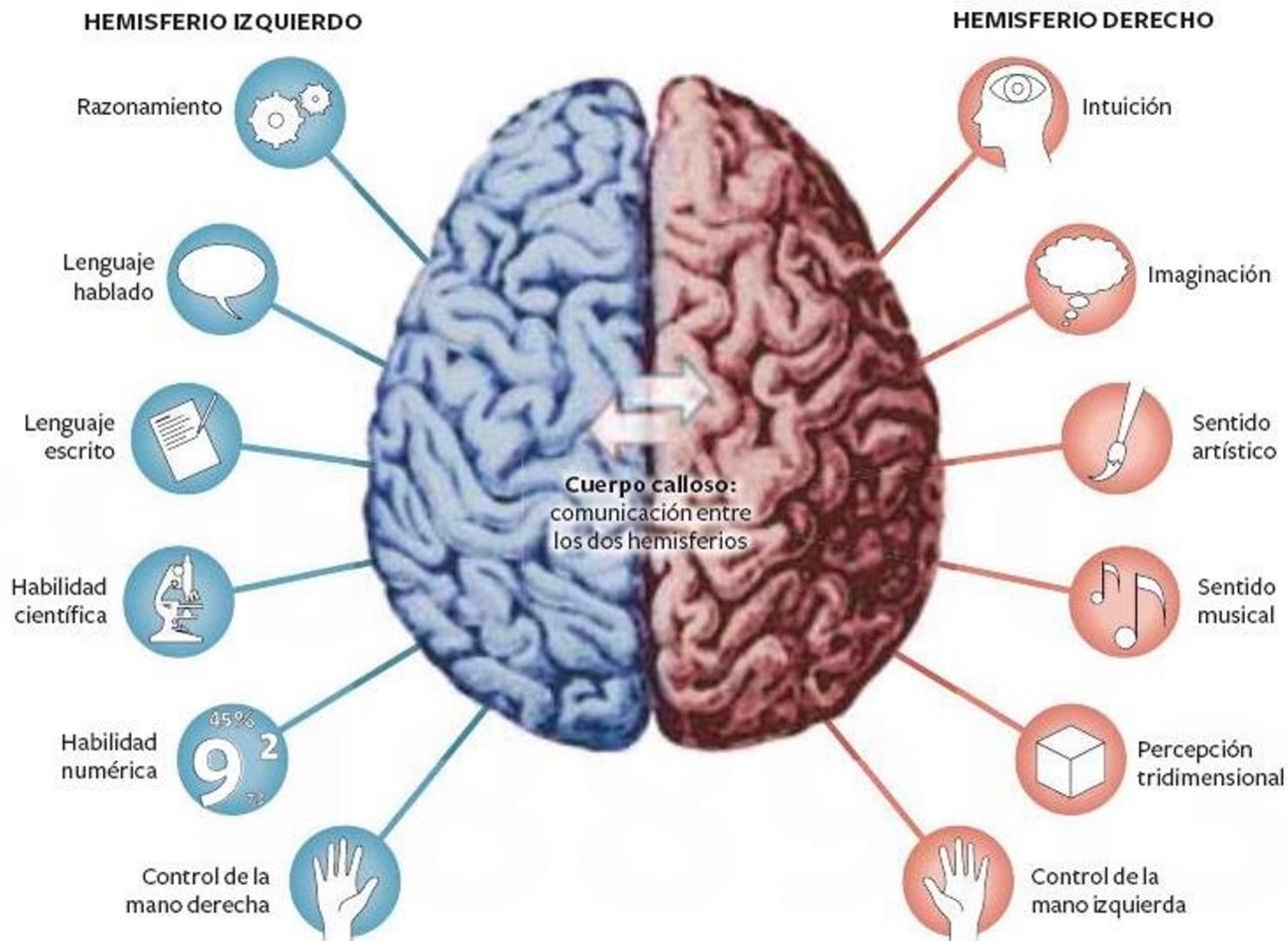
**NETFLIX**

**ROBLOX**

 **bitcoin**







# CAPACIDAD DE APRENDIZAJE



# Inteligencia artificial

Transformando la *posibilidad* en realidad a través del aprendizaje autónomo de las máquinas.

**Inteligencia Artificial (IA):** Ciencia que permite a las máquinas realizar tareas que requieren inteligencia humana. Inicia en la década de 1950.

**Alan Turing, John McCarthy**

**Machine Learning (ML):** Subcampo de la IA que permite a las máquinas aprender de datos.

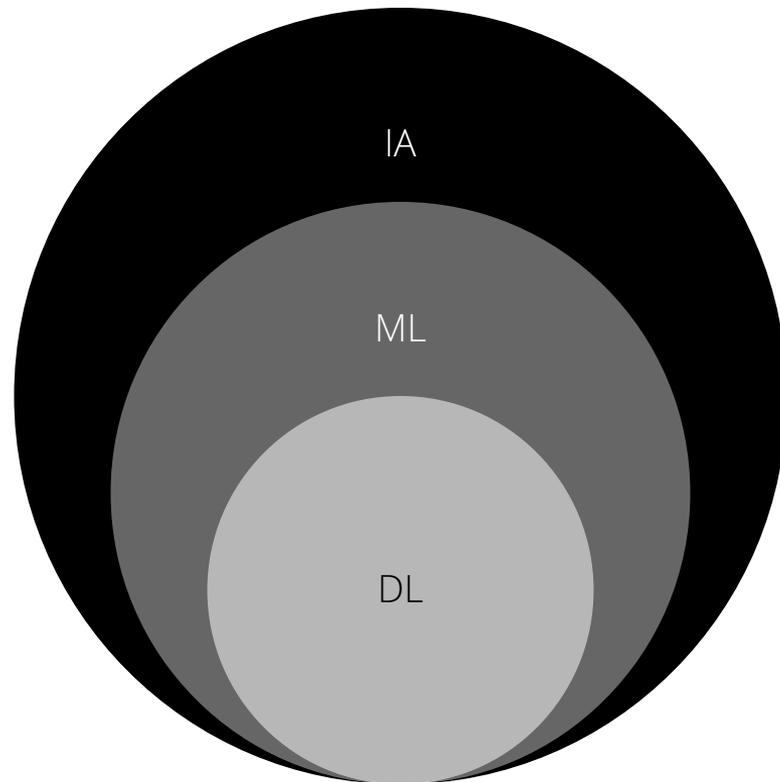
Se desarrolla en la década de 1980 como un subcampo de la IA.

**Arthur Samuel, Tom Mitchell**

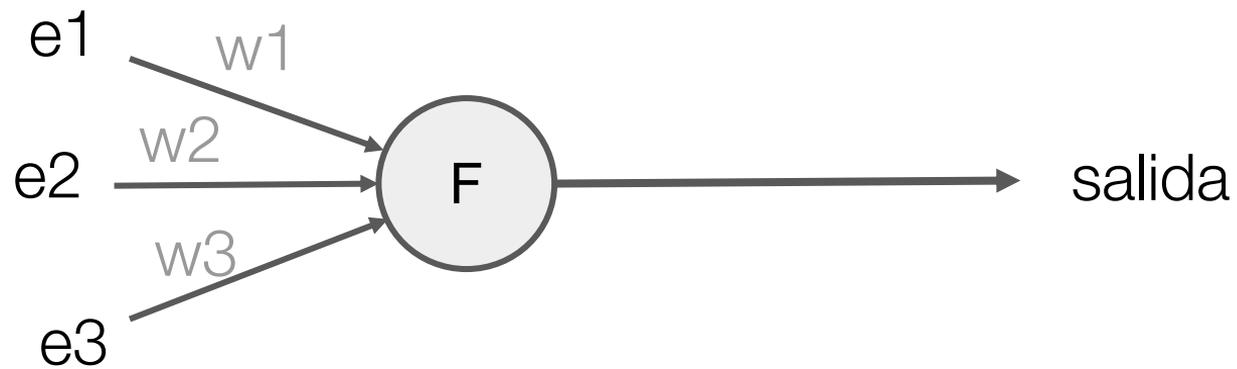
**Deep Learning (DL):** Técnica de ML que utiliza redes neuronales profundas para resolver tareas complejas.

Gana relevancia en la década de 2010, aunque sus raíces se remontan a algoritmos más antiguos.

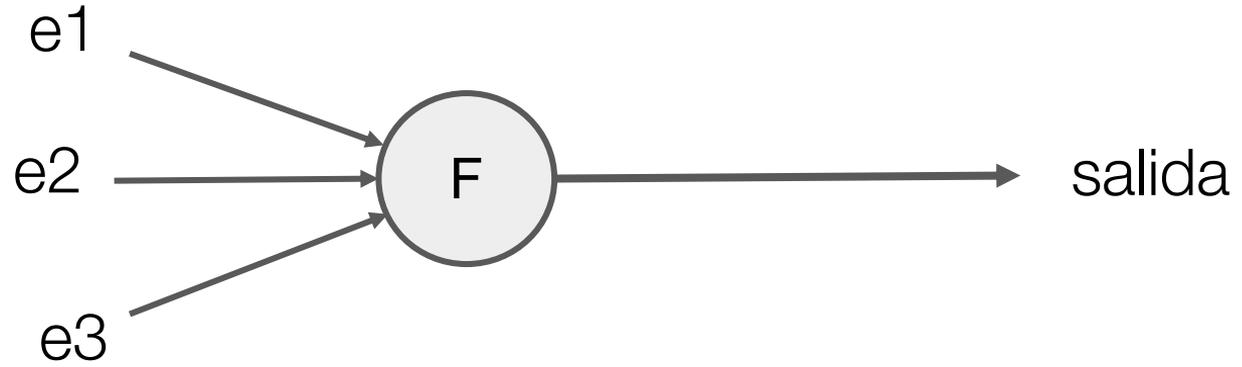
**Geoffrey Hinton, Yann LeCun**











TENER EL DINERO SUFICIENTE

-> SI

/ NO

TU PAREJA QUIERE IR DE VIAJE

-> SI

/ NO

EL LUGAR TIENE CLIMA AGRADABLE

-> SI / NO

SI

= 1

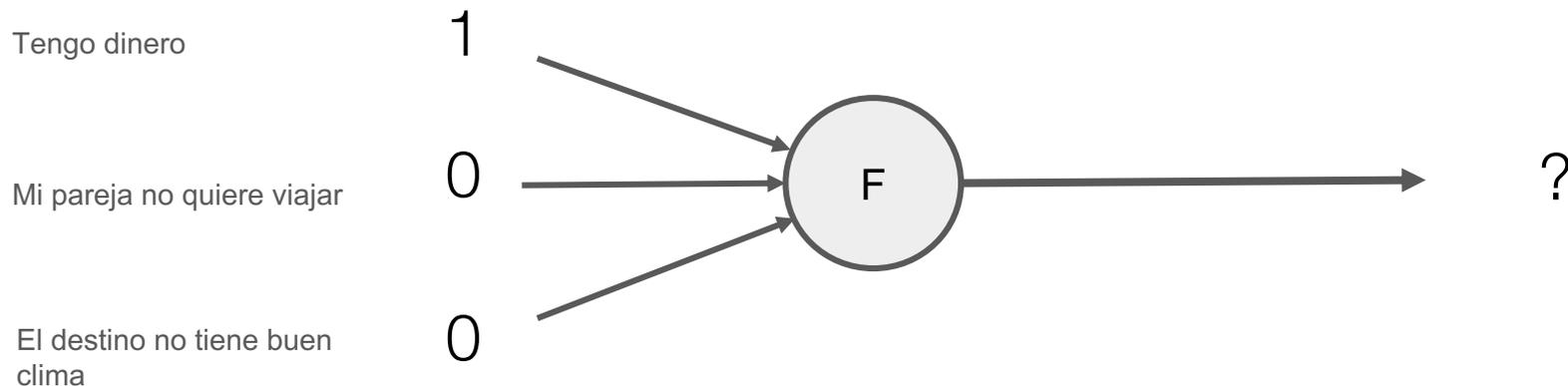
NO

= 0

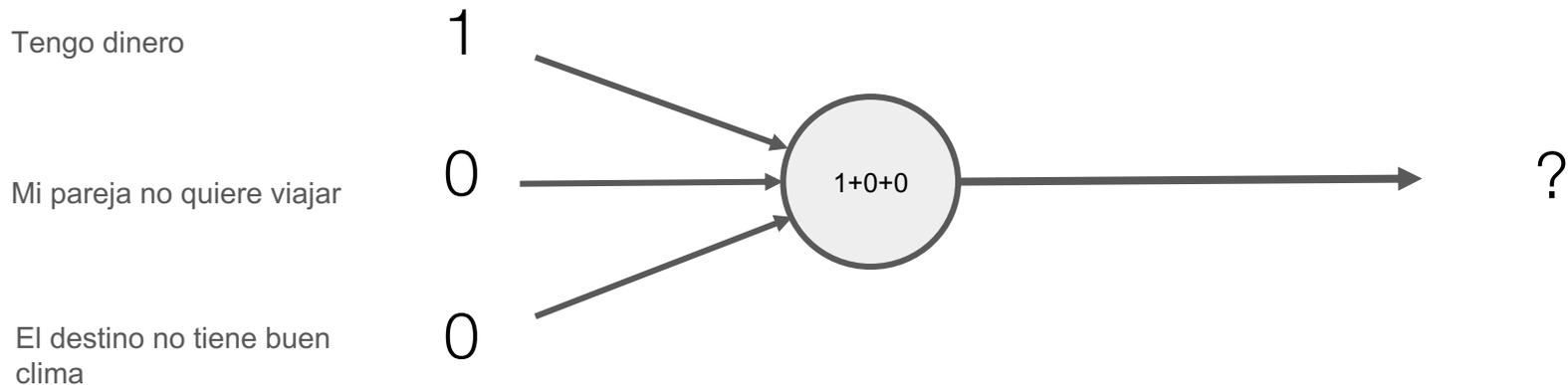
# Función de activación

**número > 1**

# Cómo decidir si viajar o no



# Cómo decidir si viajar o no



# Cómo decidir si viajar o no

Tengo dinero

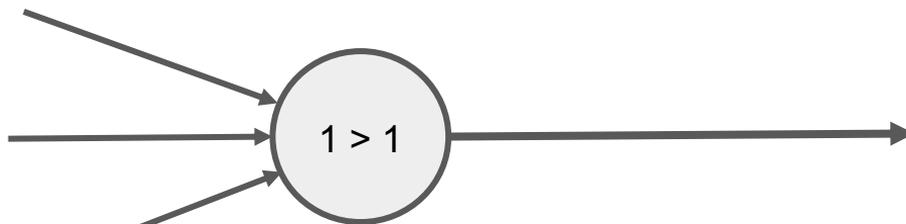
1

Mi pareja no quiere viajar

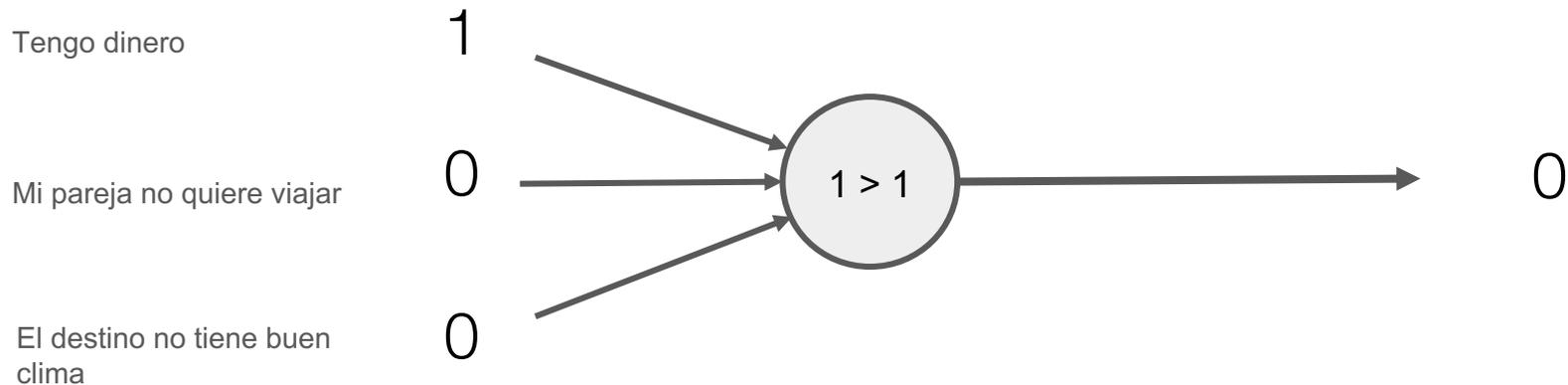
0

El destino no tiene buen  
clima

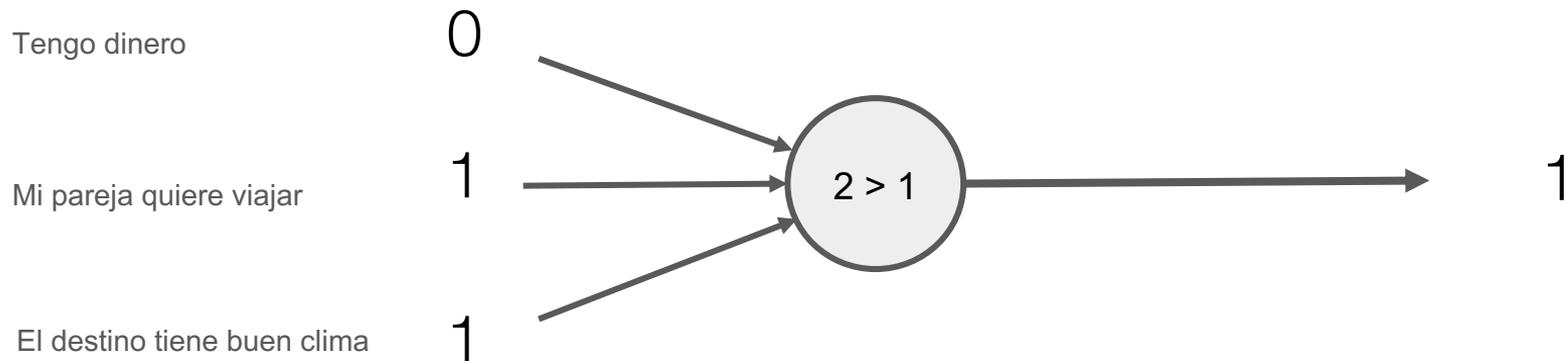
0



# Cómo decidir si viajar o no



## Cómo decidir si viajar o no





# Cómo decidir si viajar o no

¿Tengo dinero?

?

2

¿Mi pareja quiere viajar?

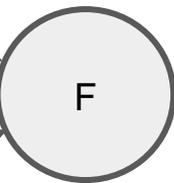
?

1

¿El destino tiene buen clima?

?

1

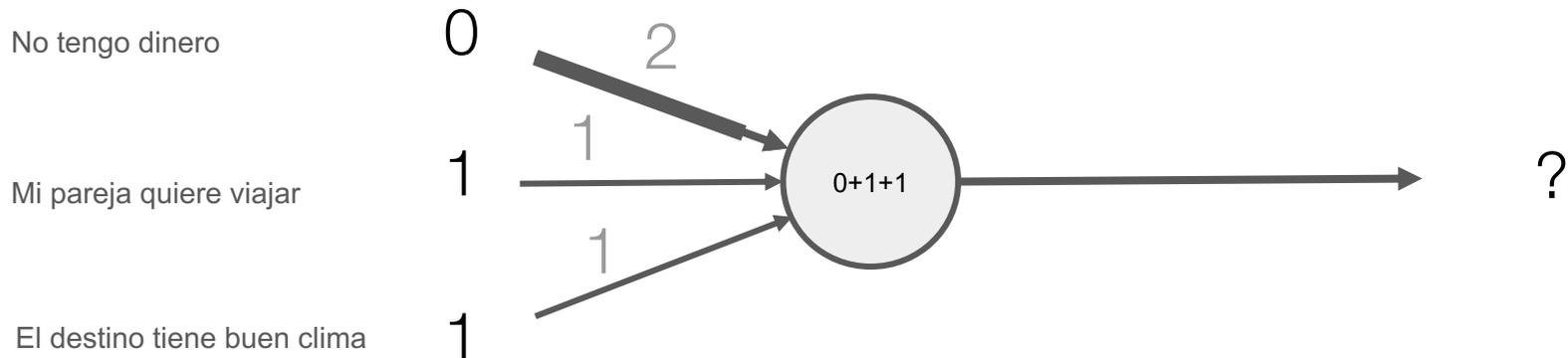


?

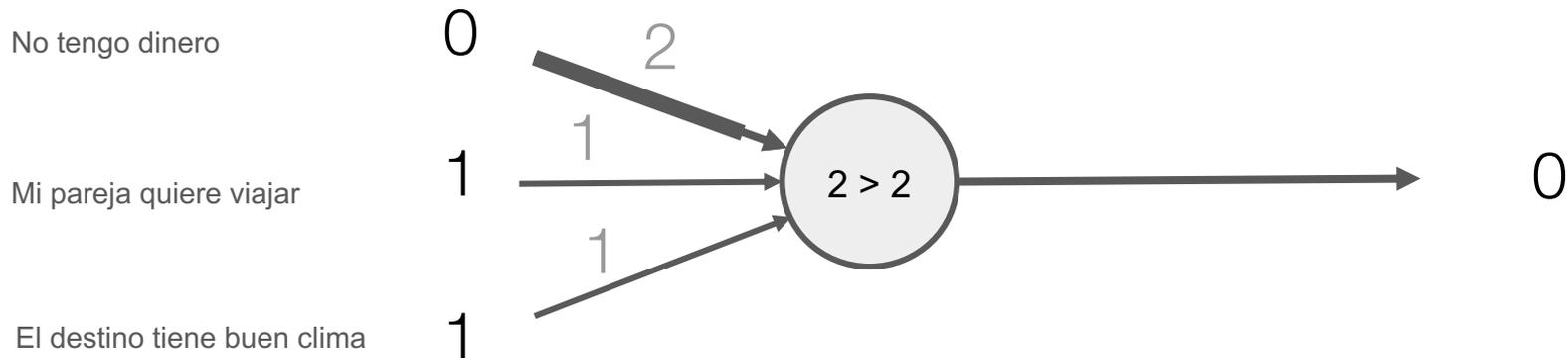
# Función de activación

**número > 2**

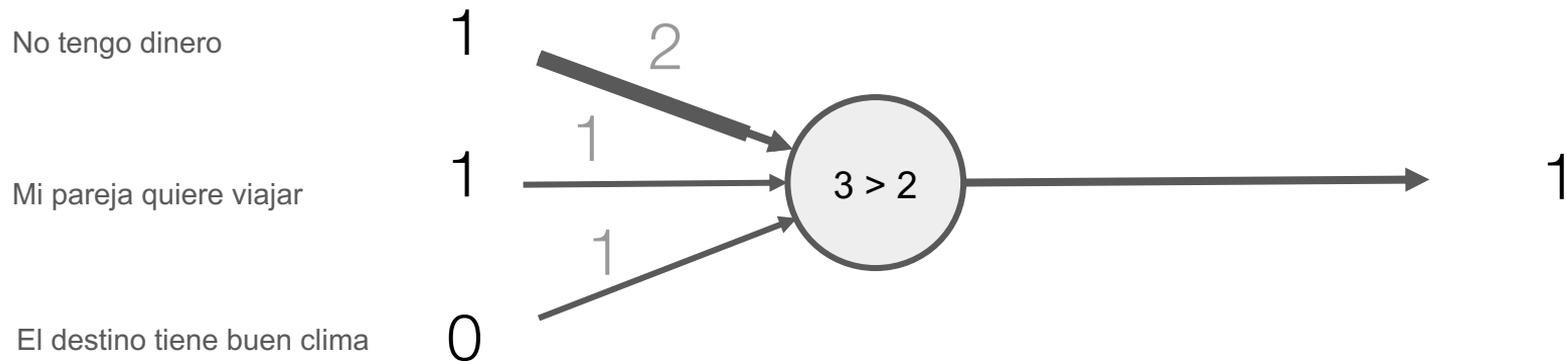
# Cómo decidir si viajar o no

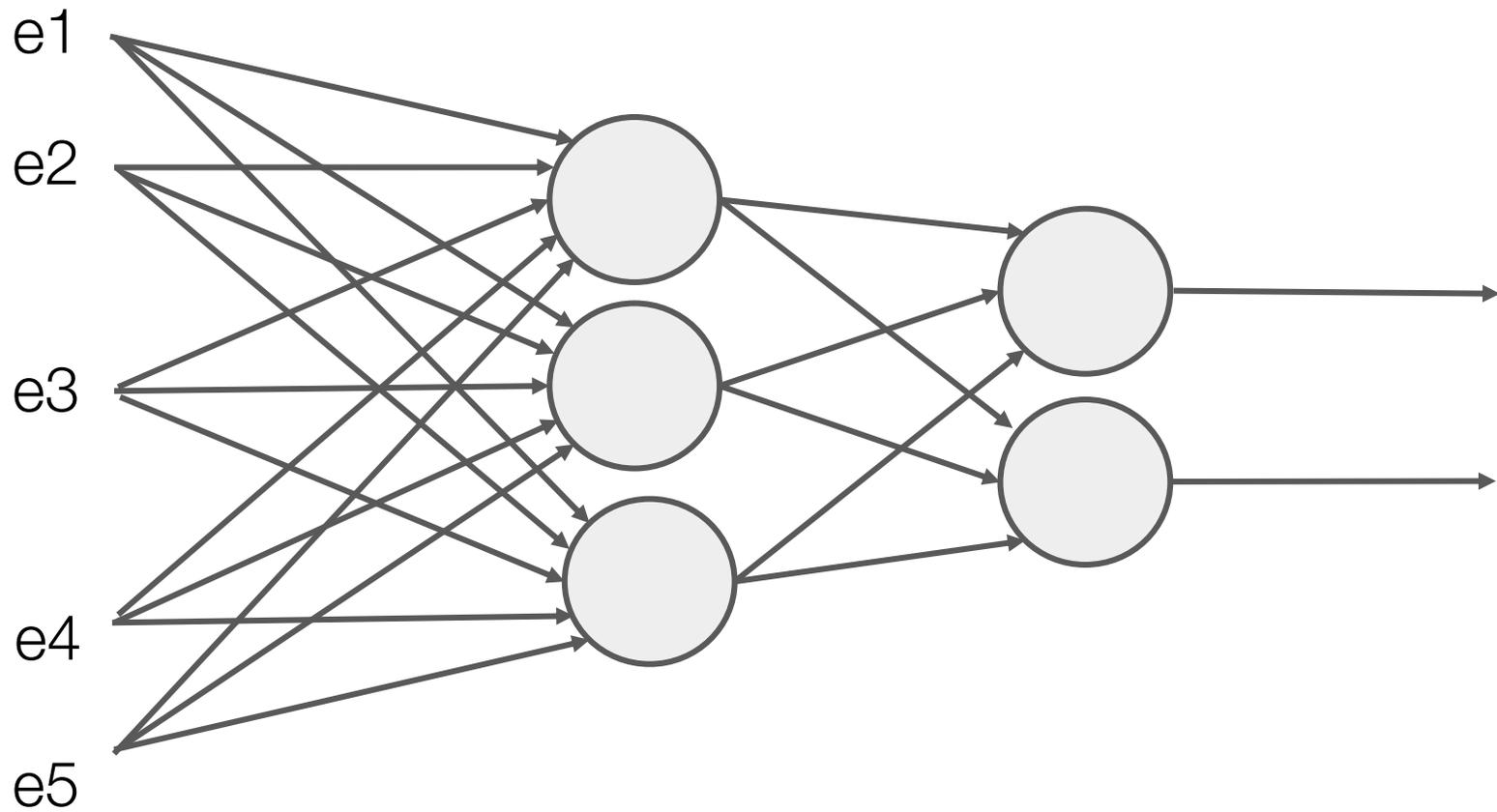


## Cómo decidir si viajar o no

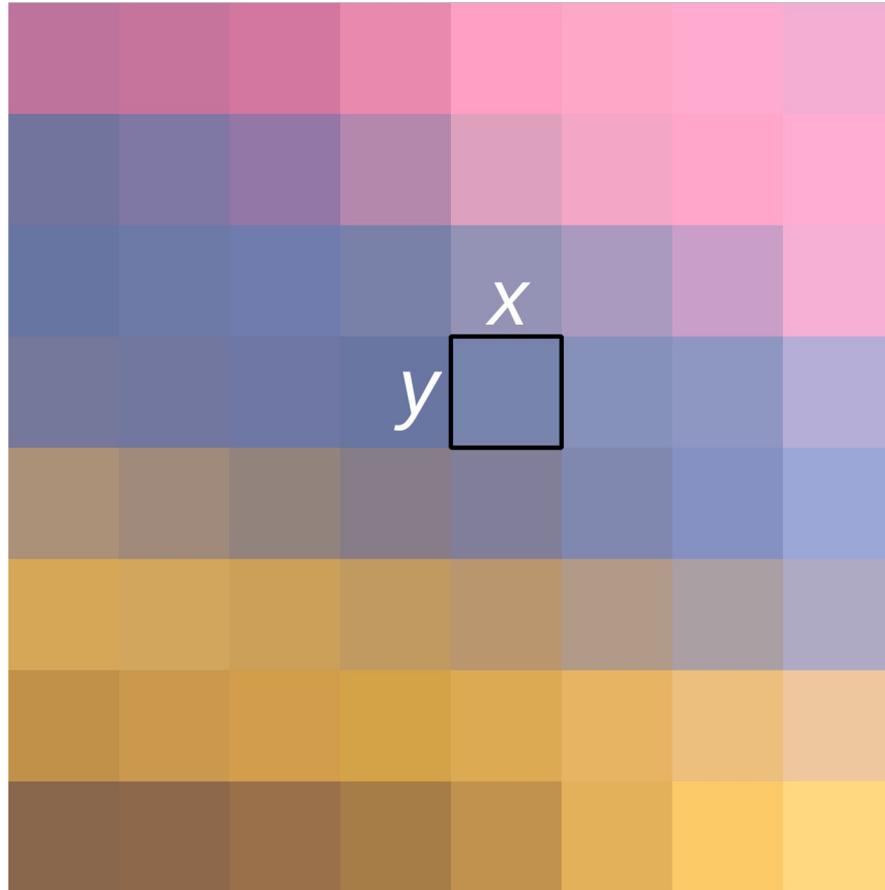


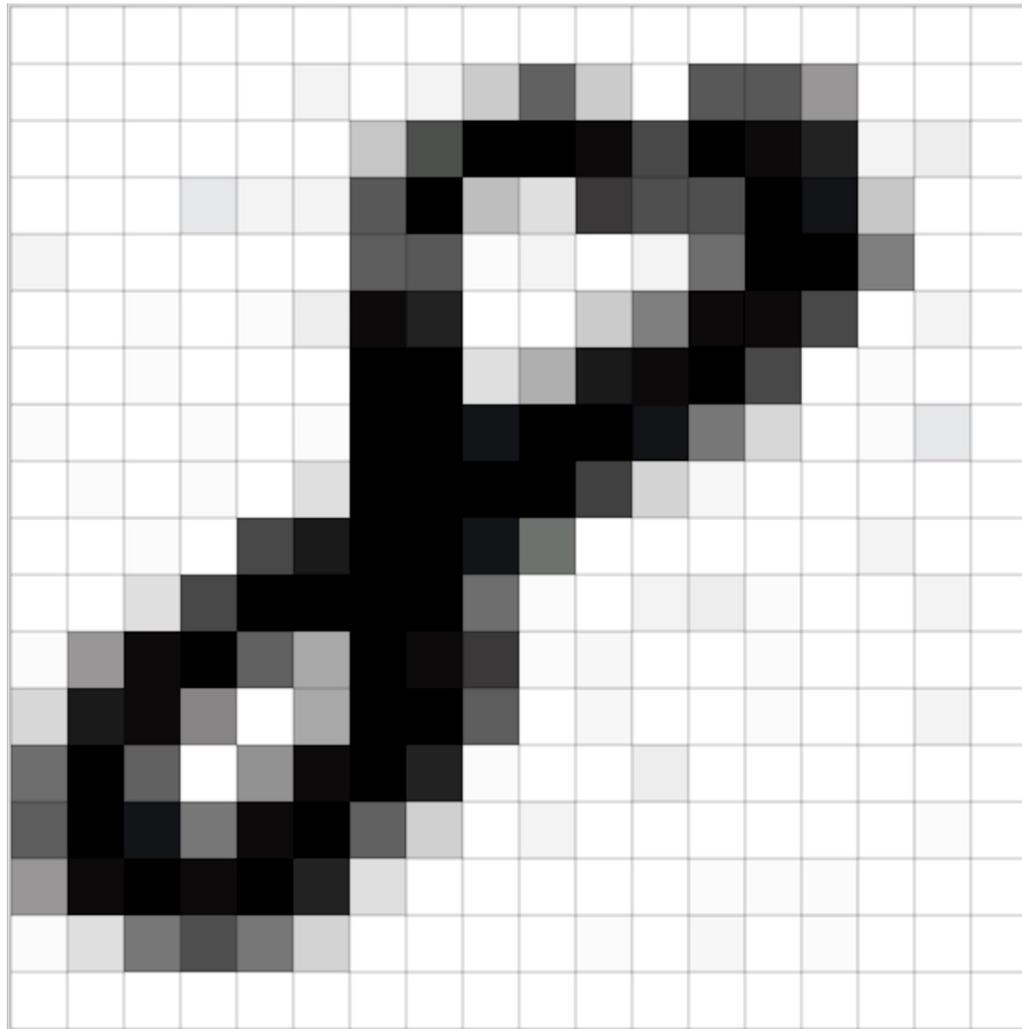
## Cómo decidir si viajar o no

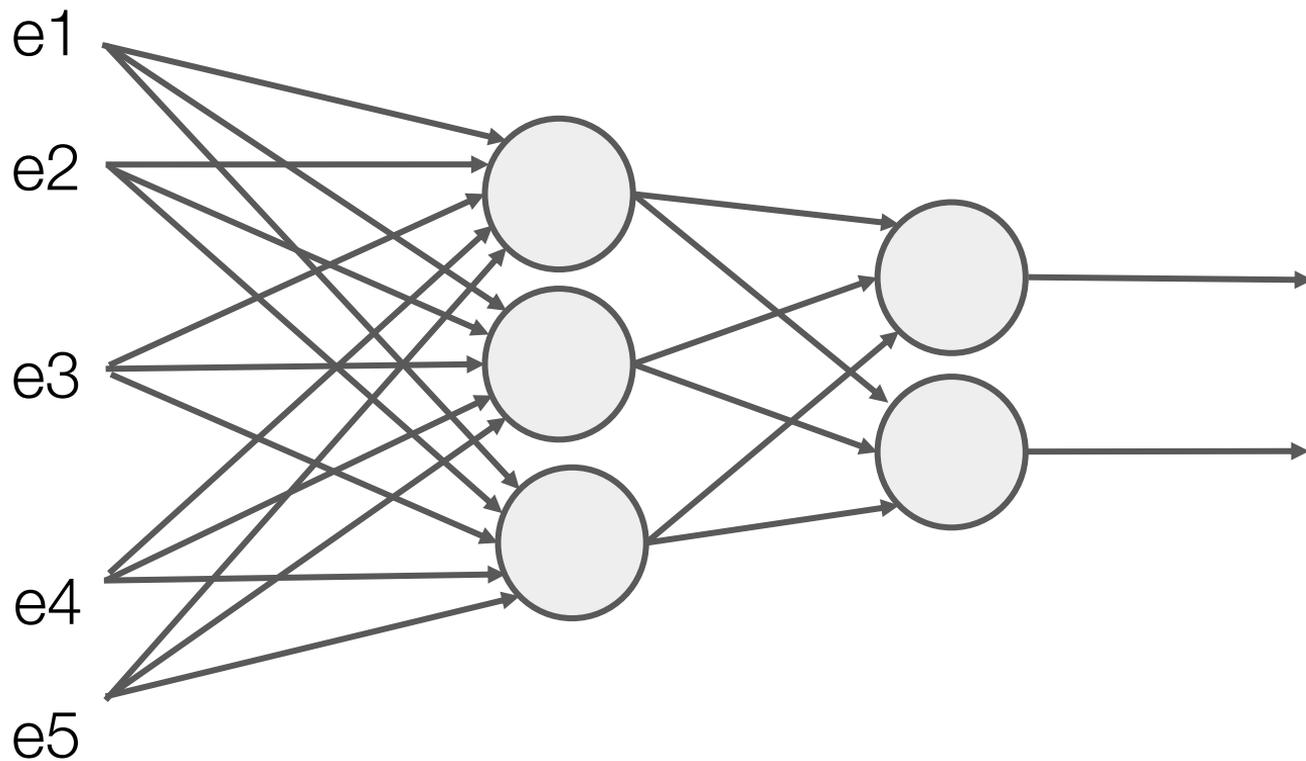
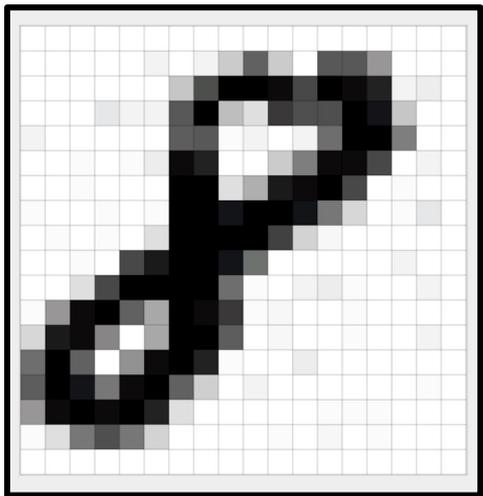






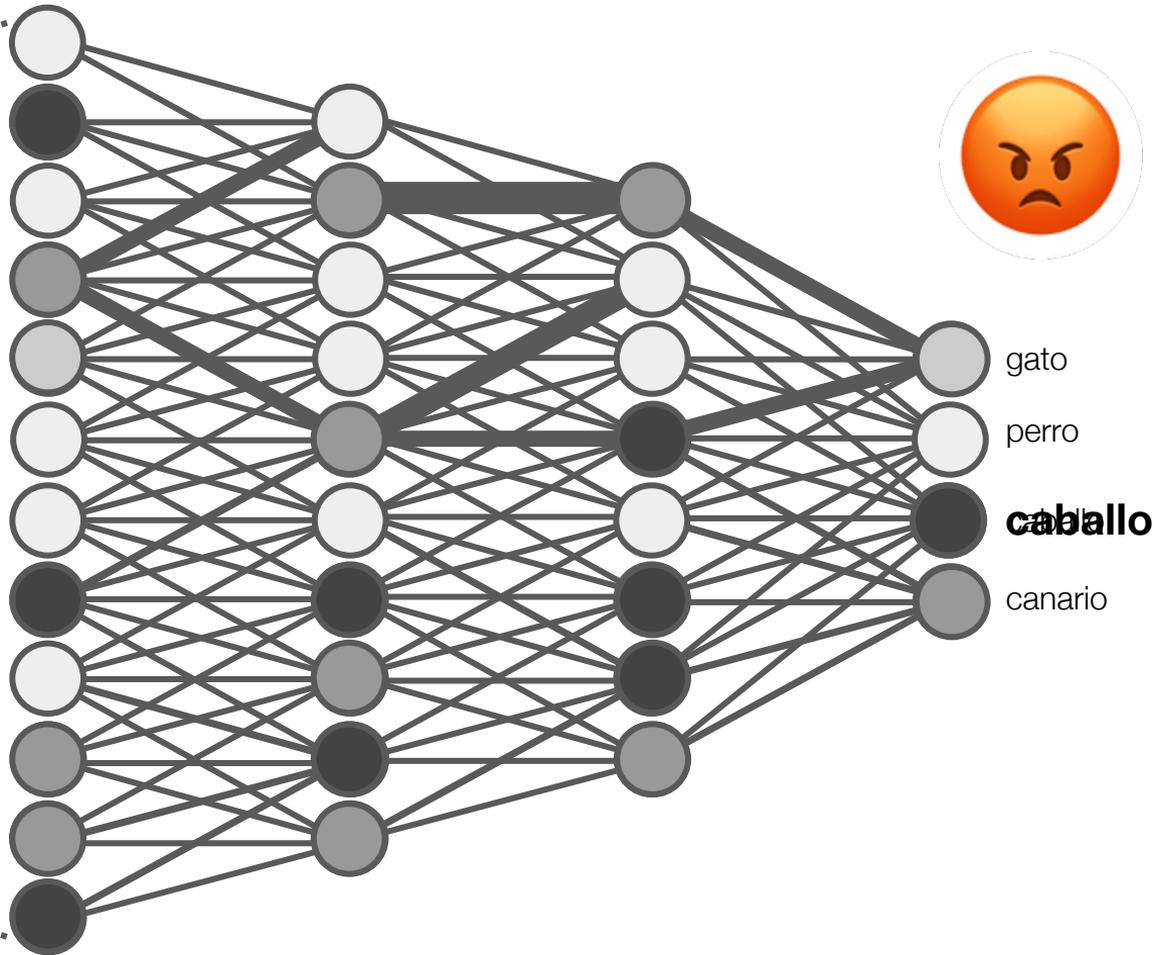






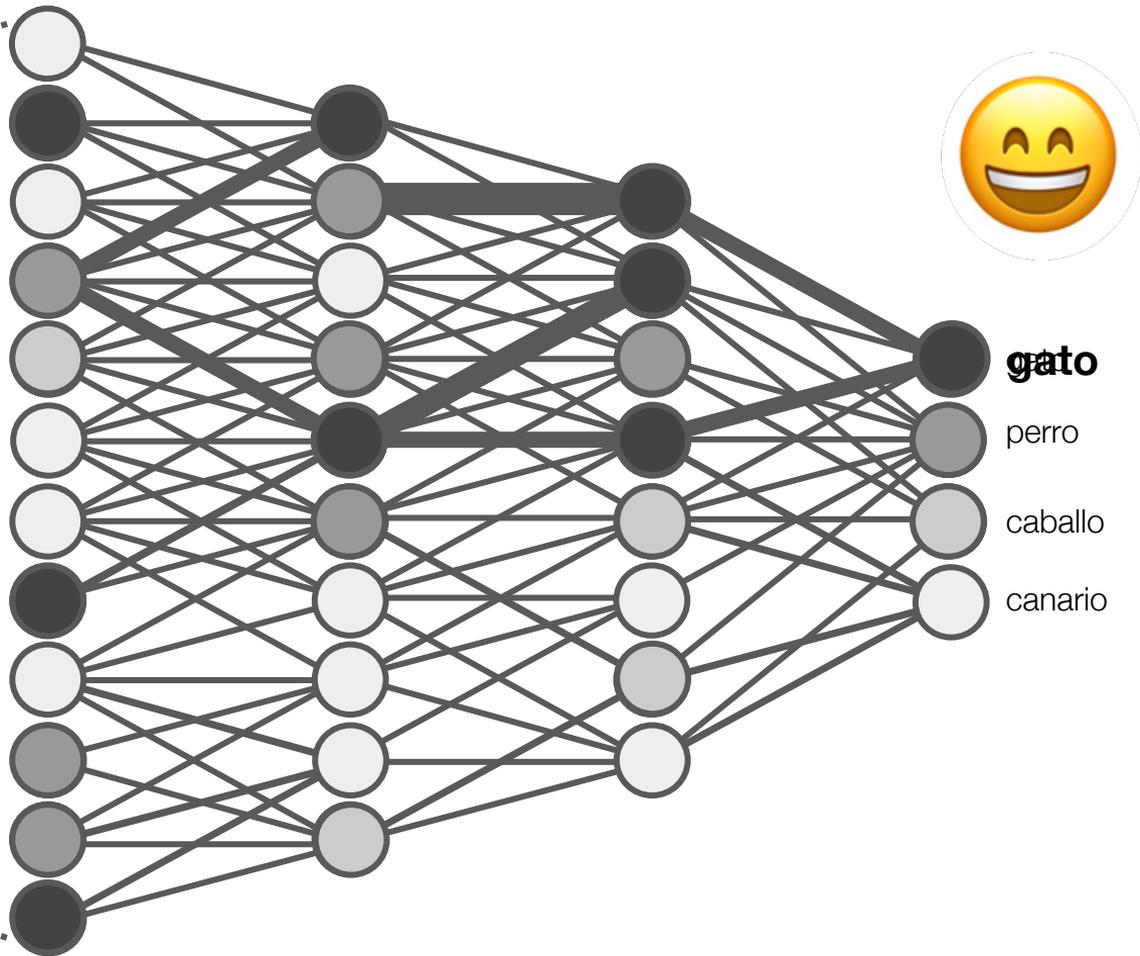


gato



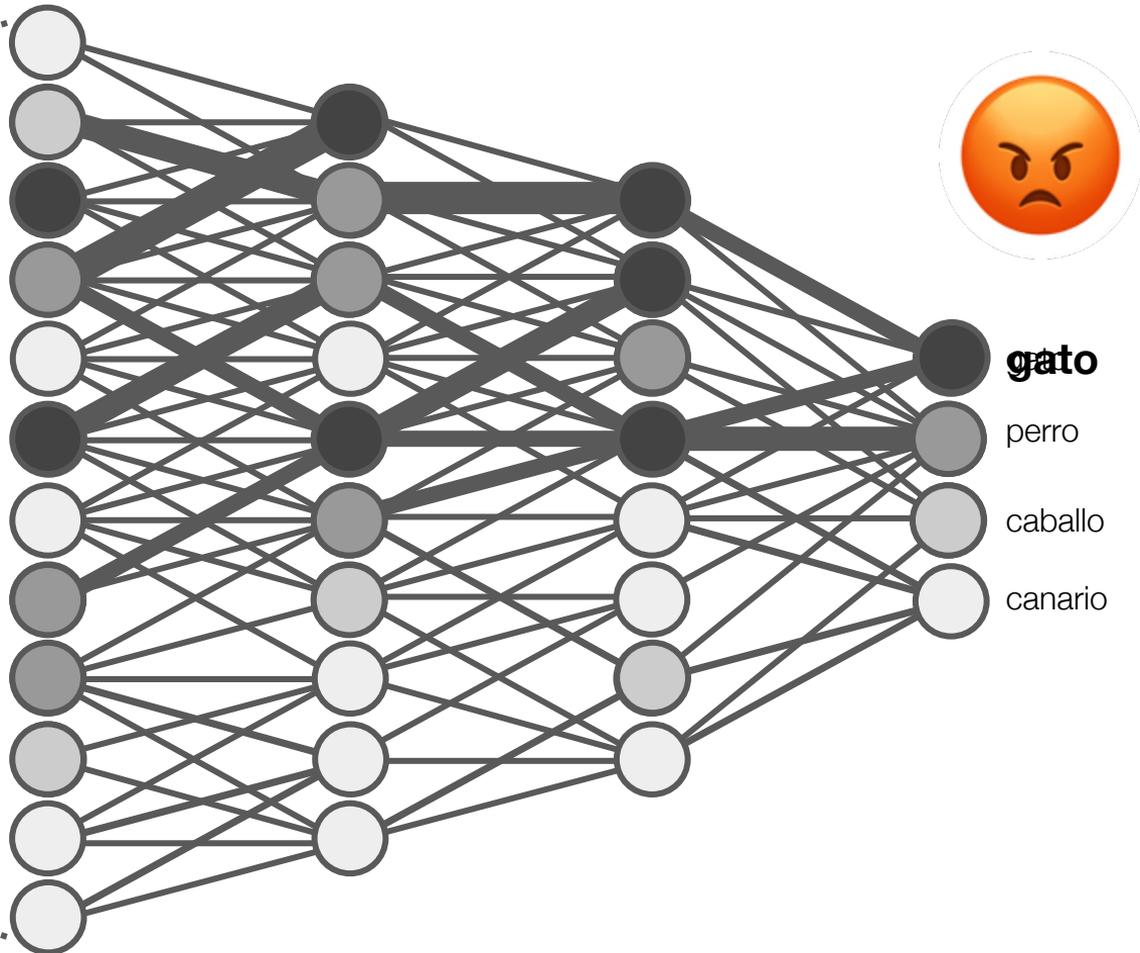


**gato**



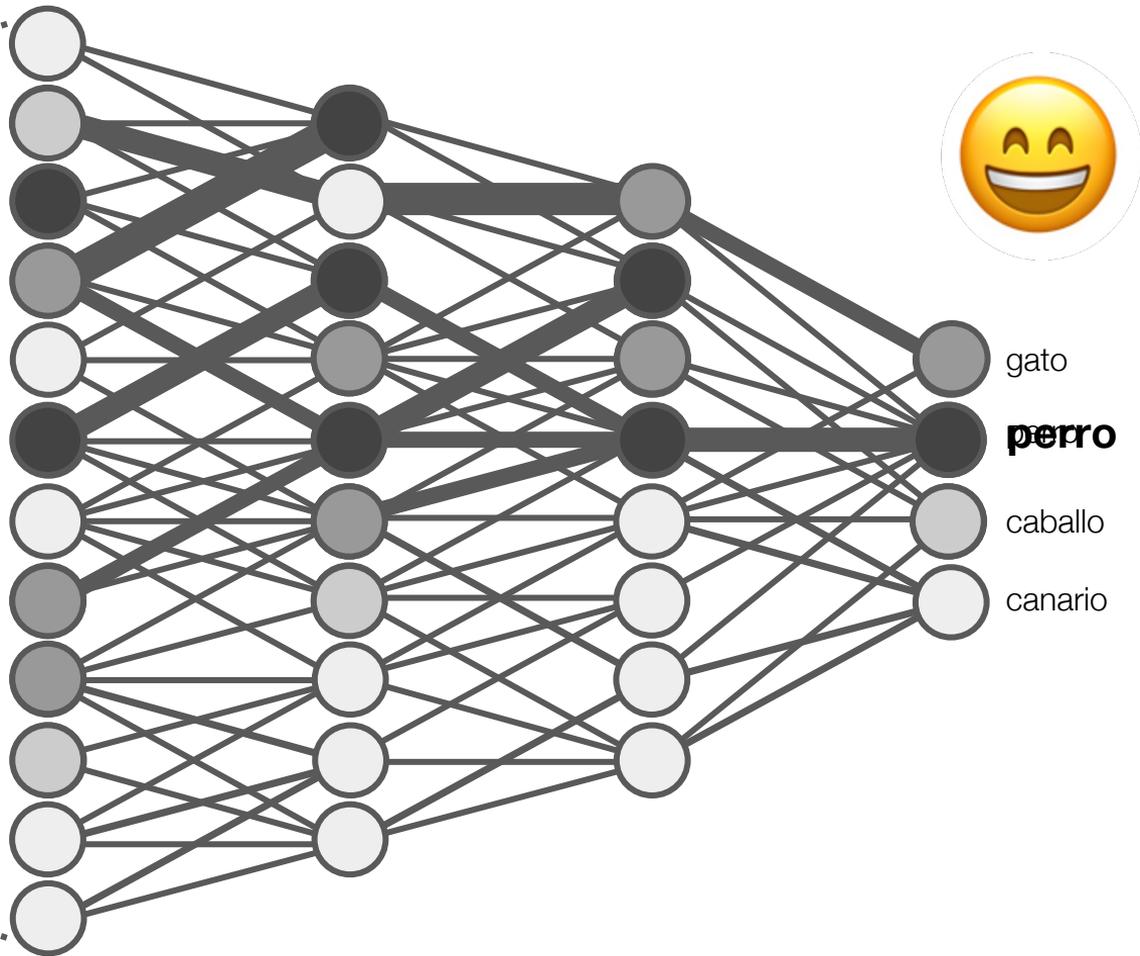


perro



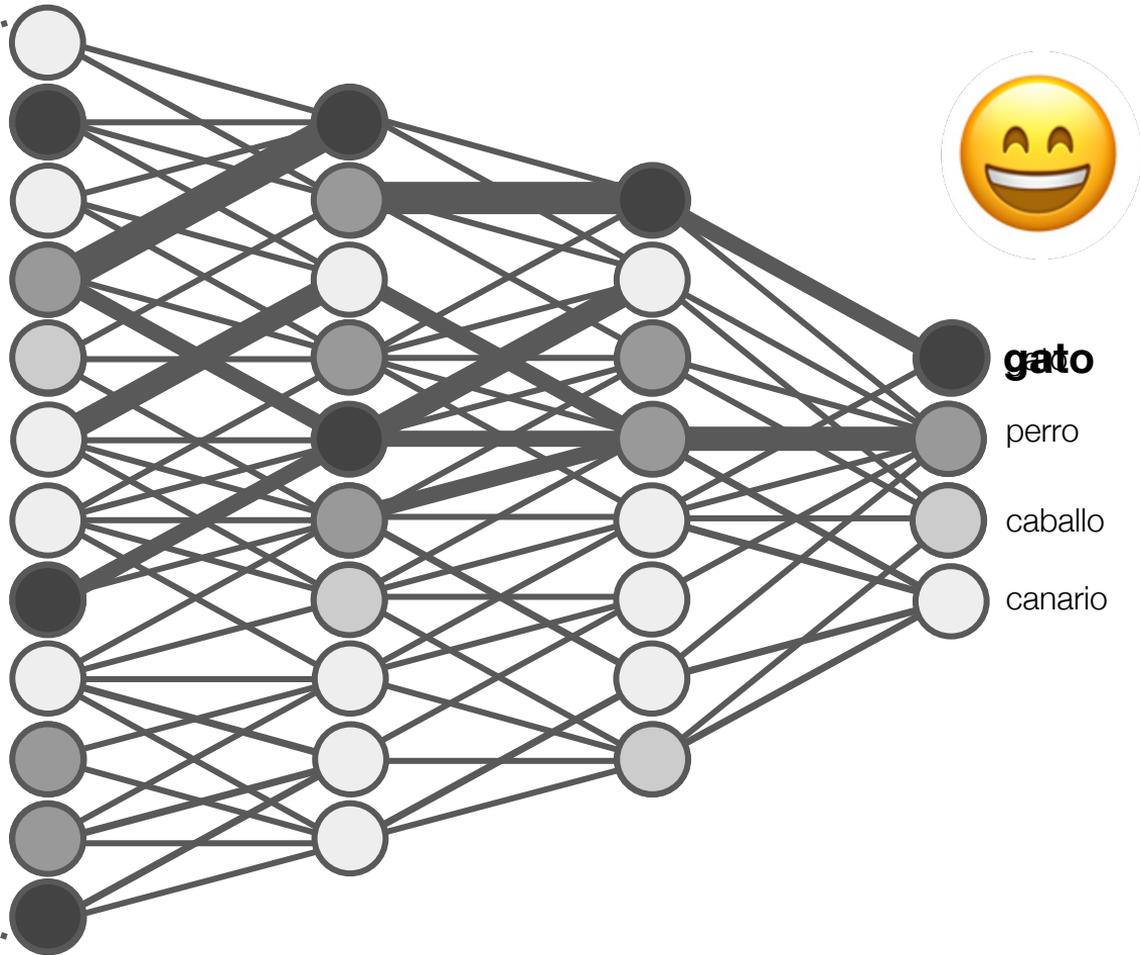


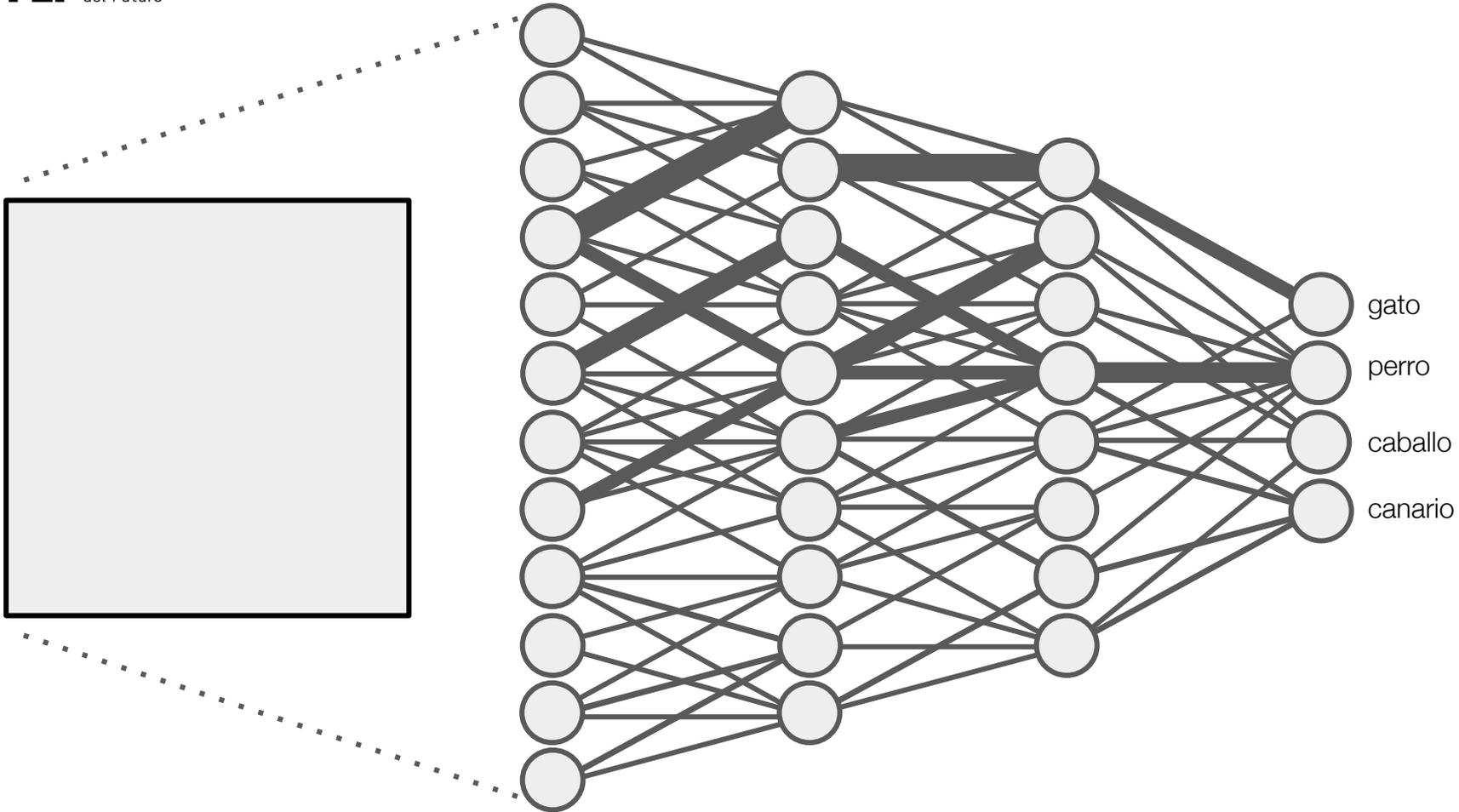
perro





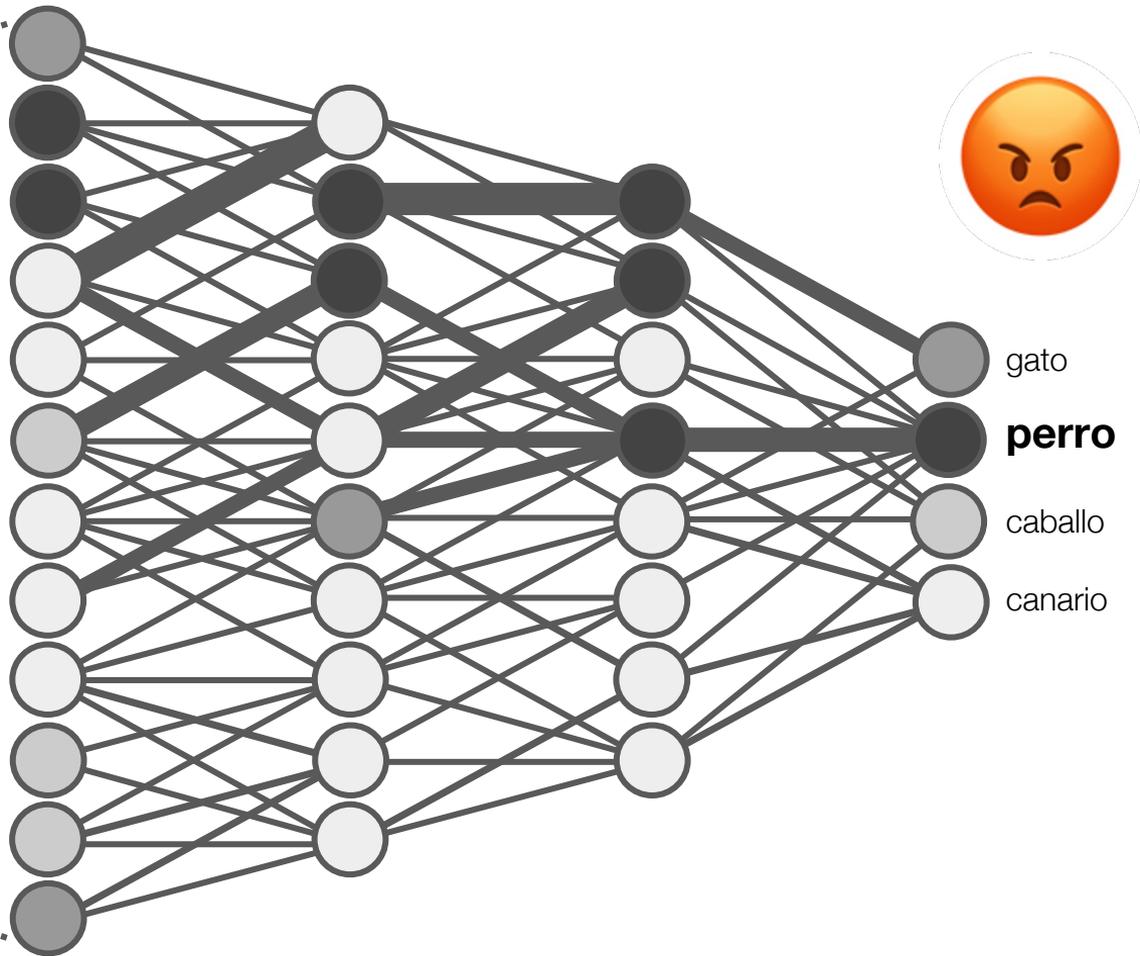
**gato**





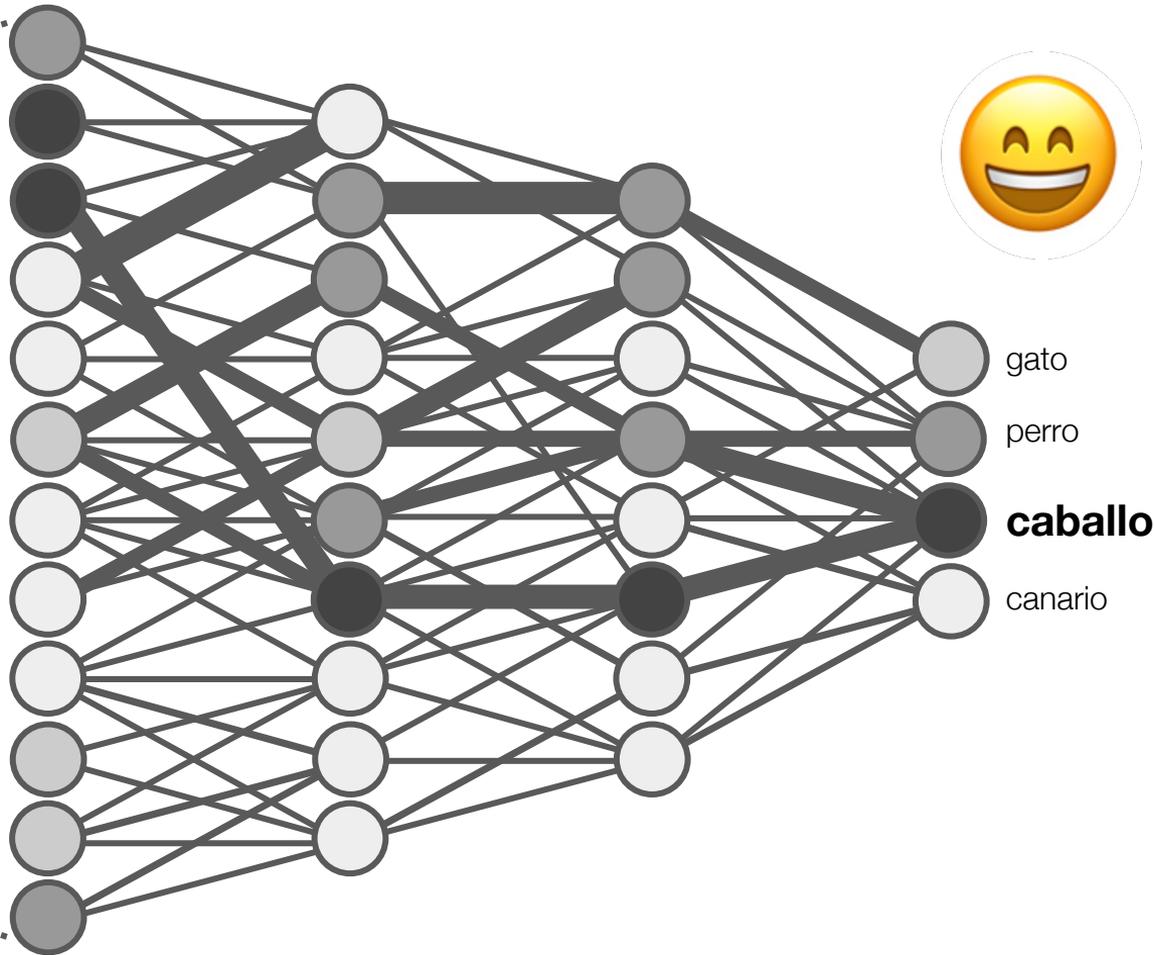


**caballo**



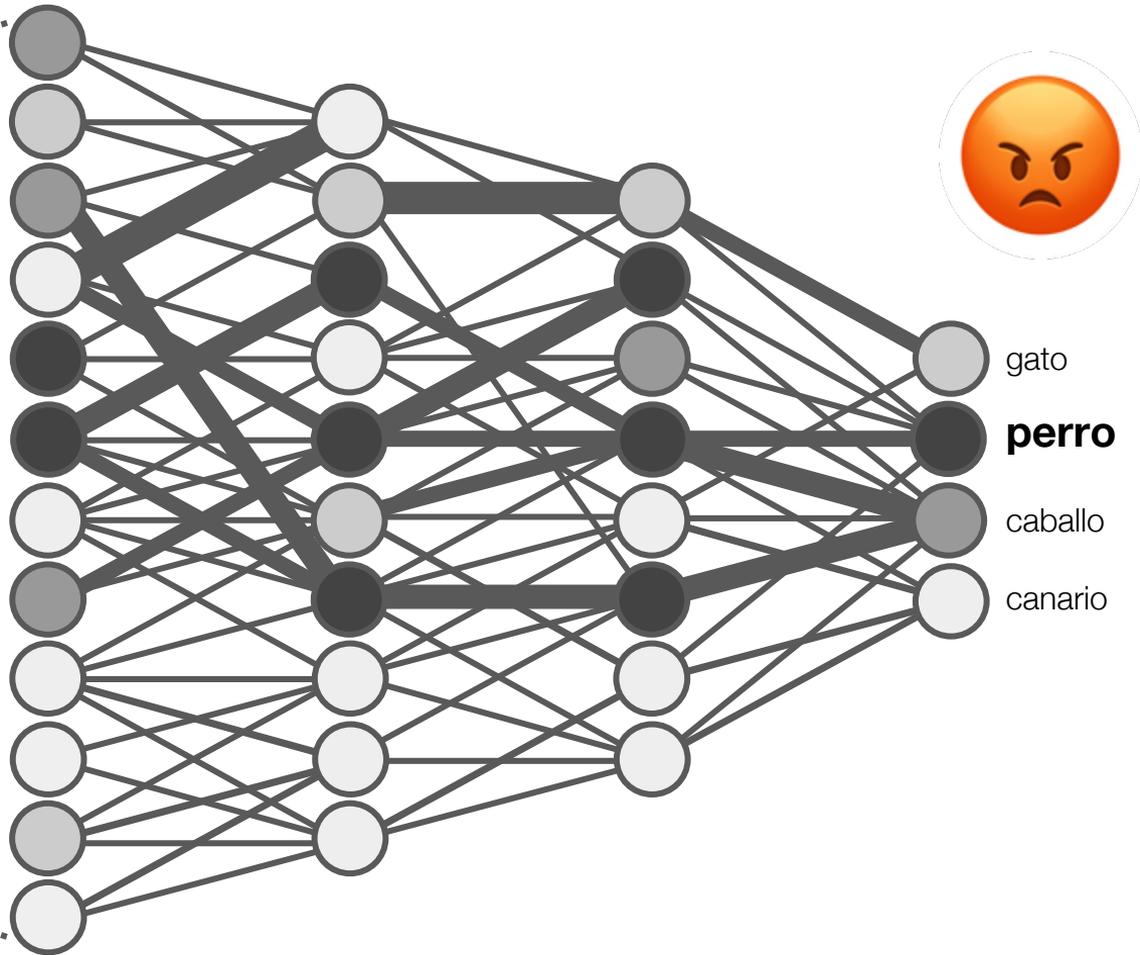


**caballo**



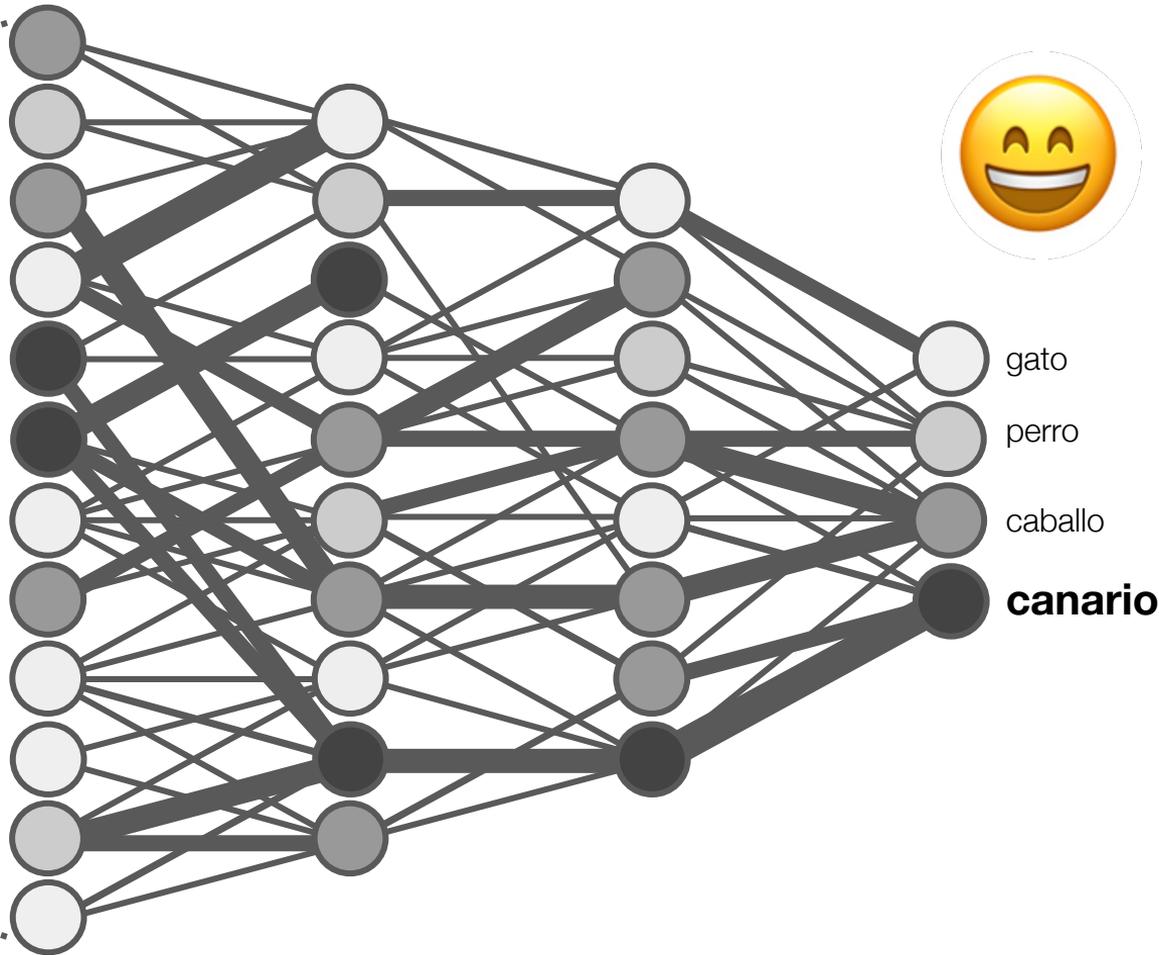


canario



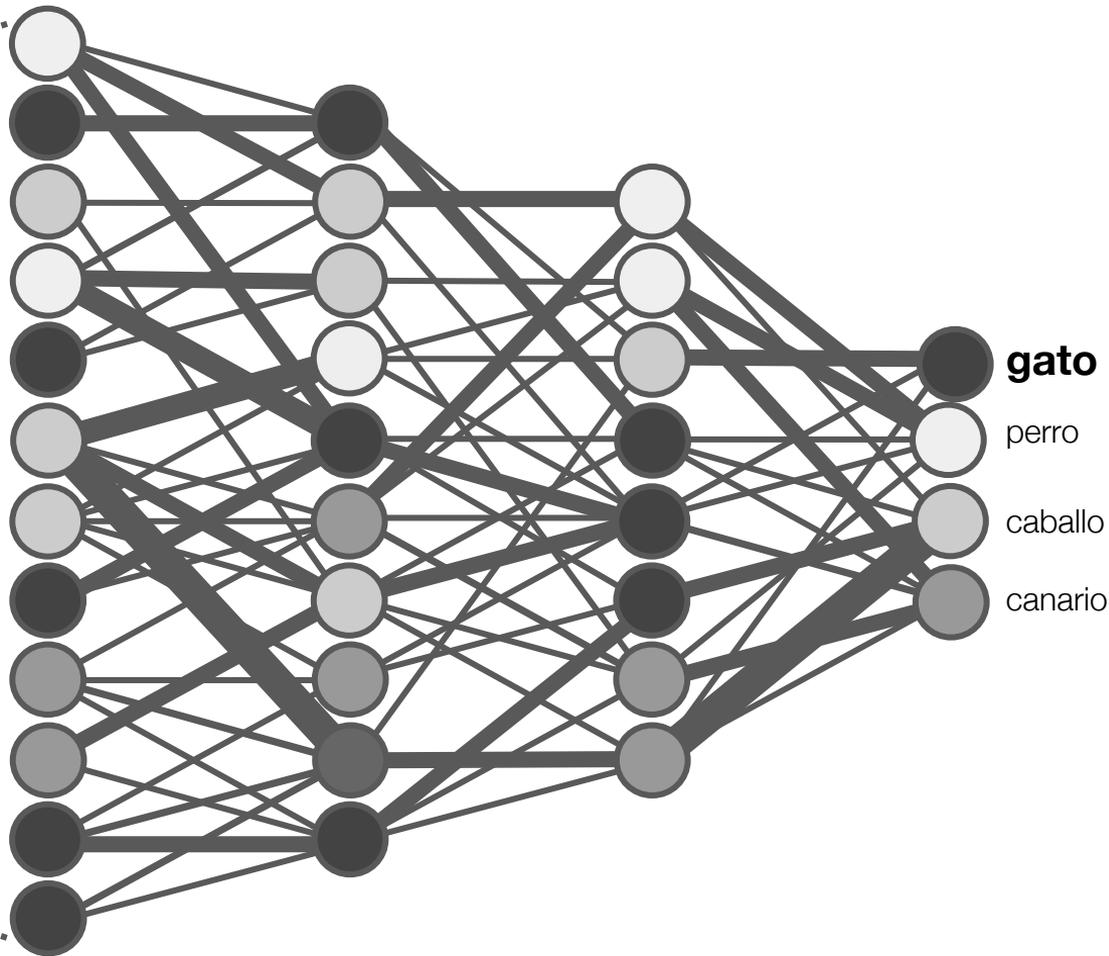


canario



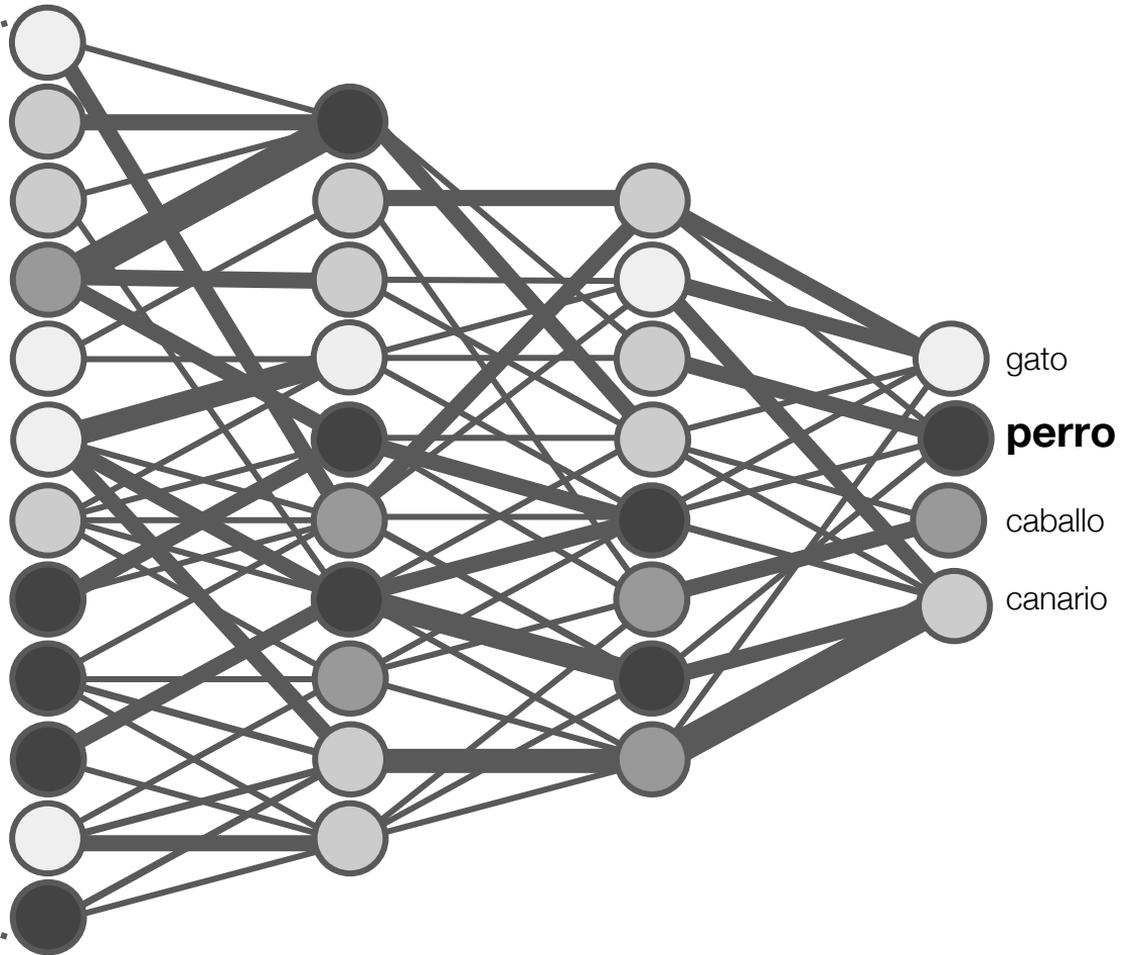


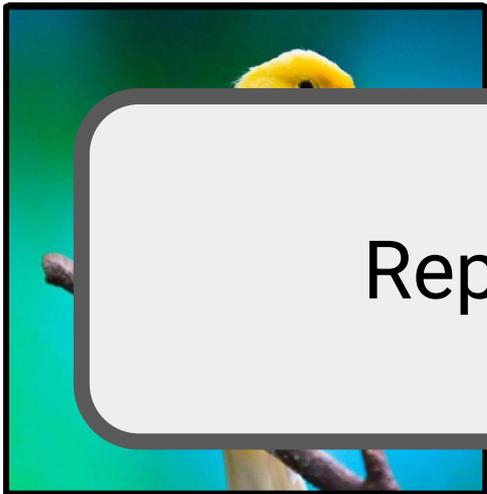
**gato**





perro





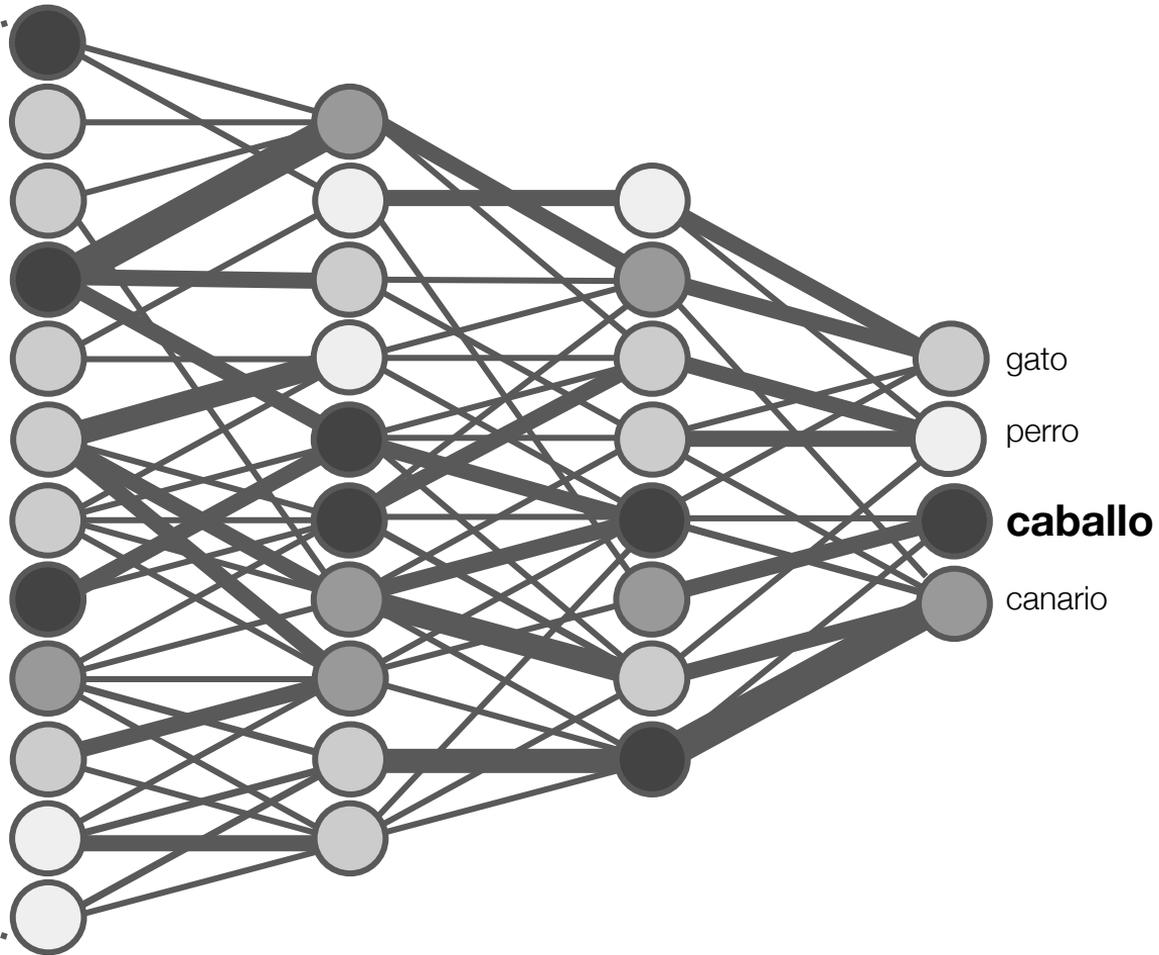
canario

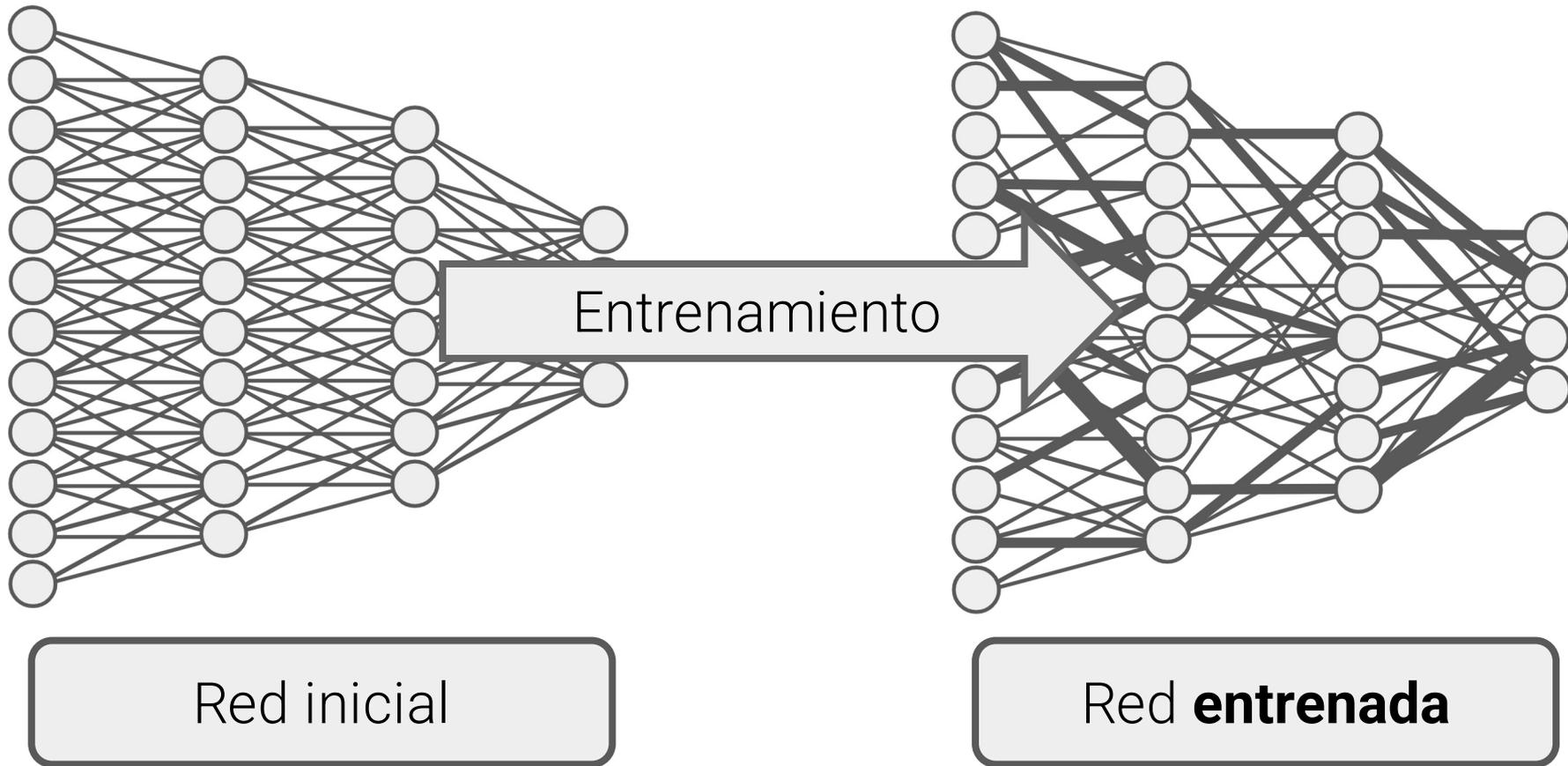


Repetimos millones de veces...



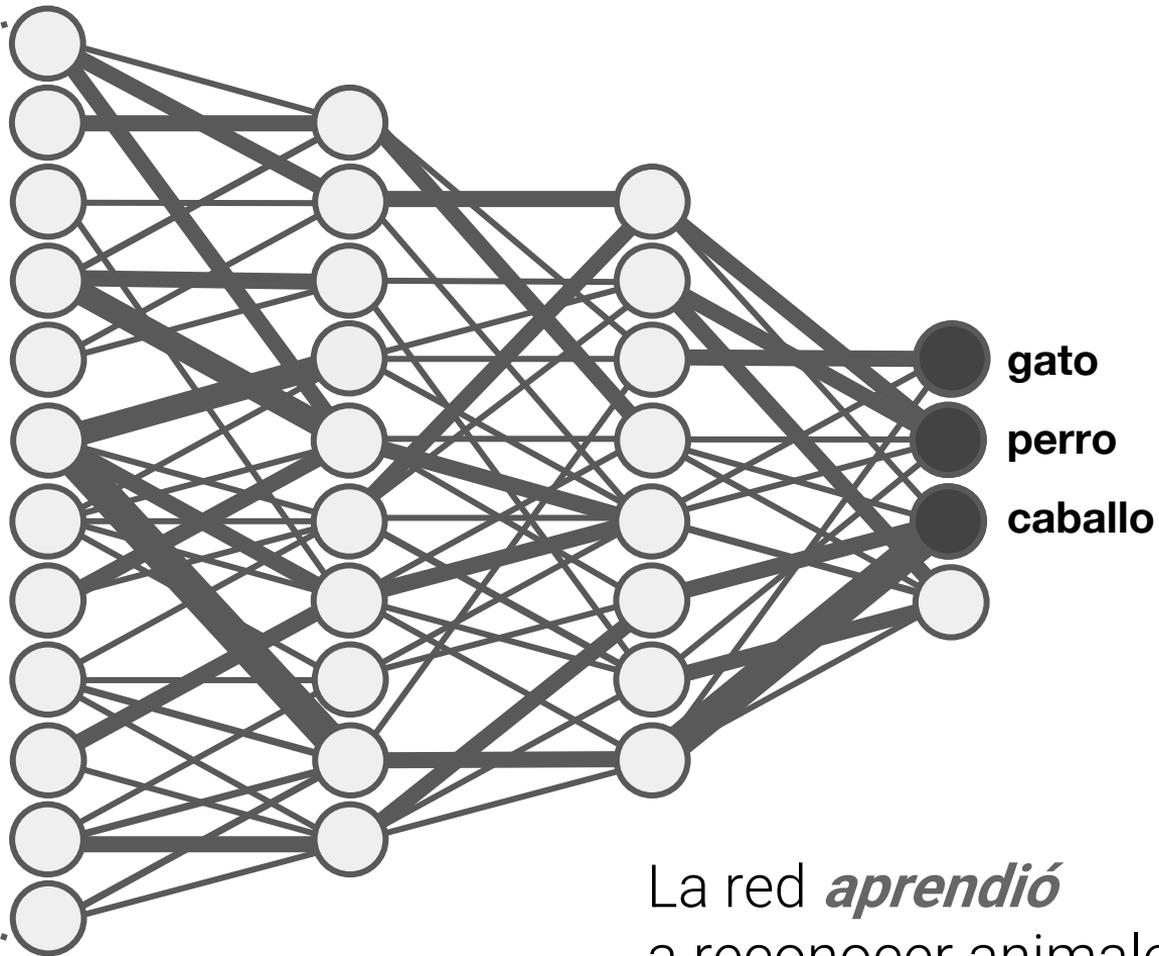
**caballo**





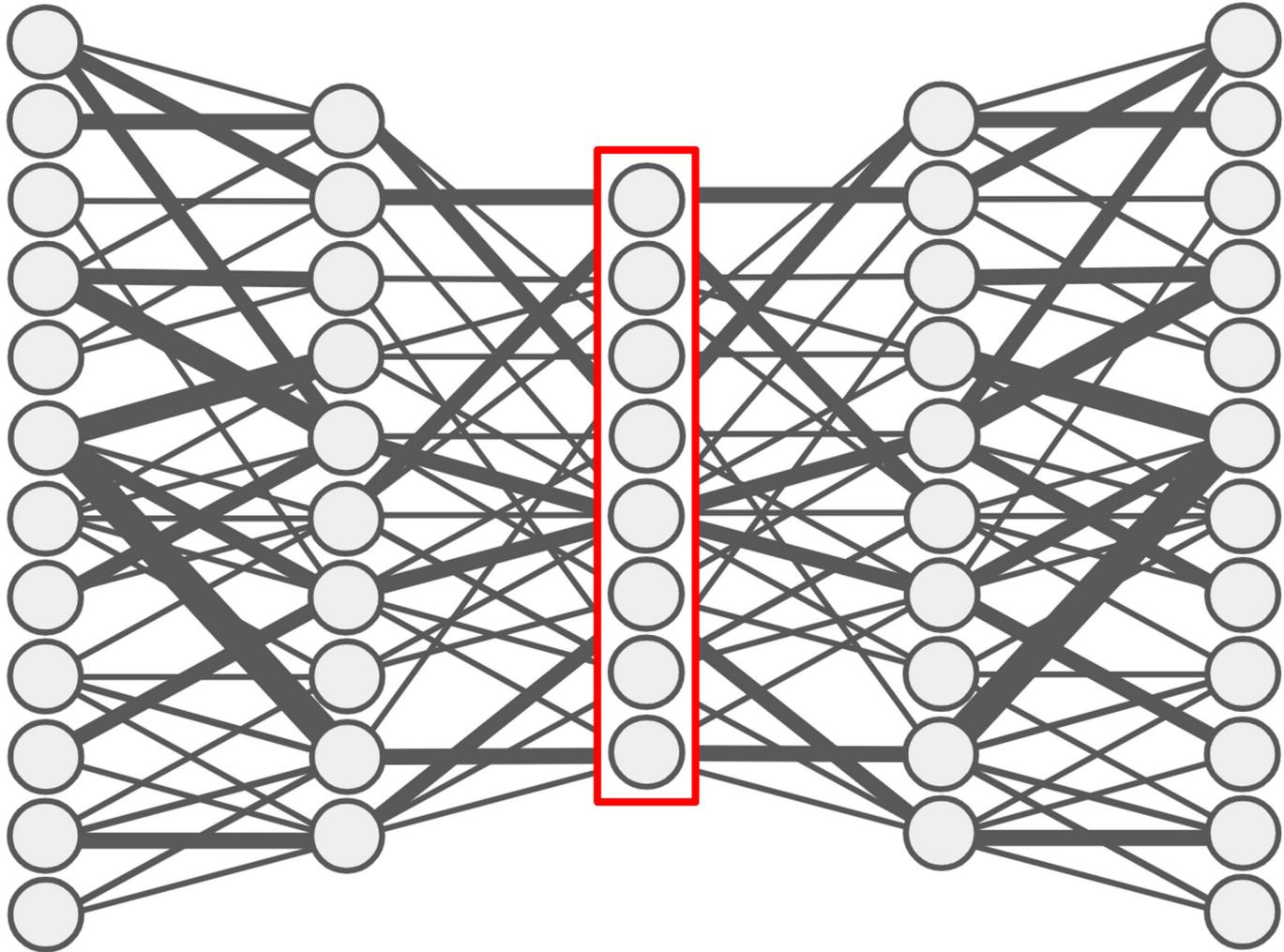


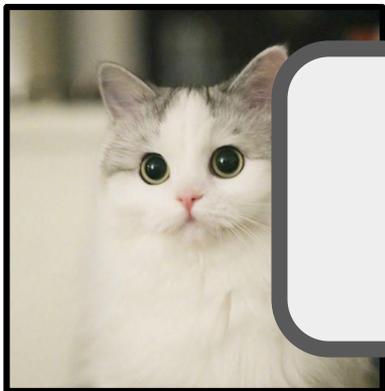
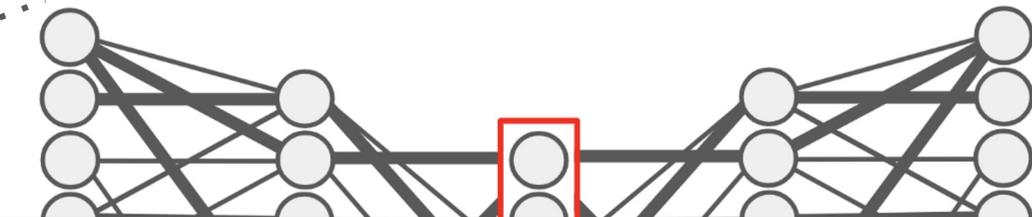
Imágenes no vistas  
durante el entrenamiento



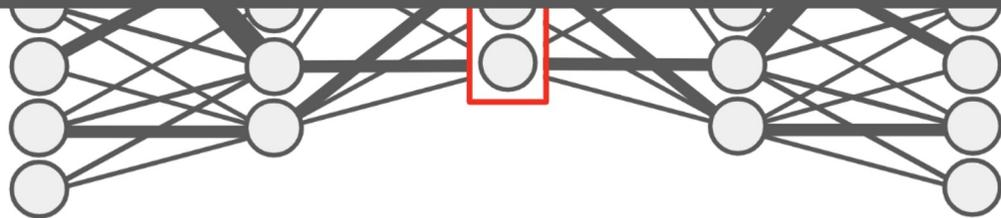
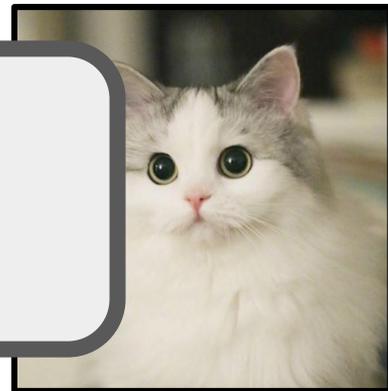
La red *aprendió*  
a reconocer animales

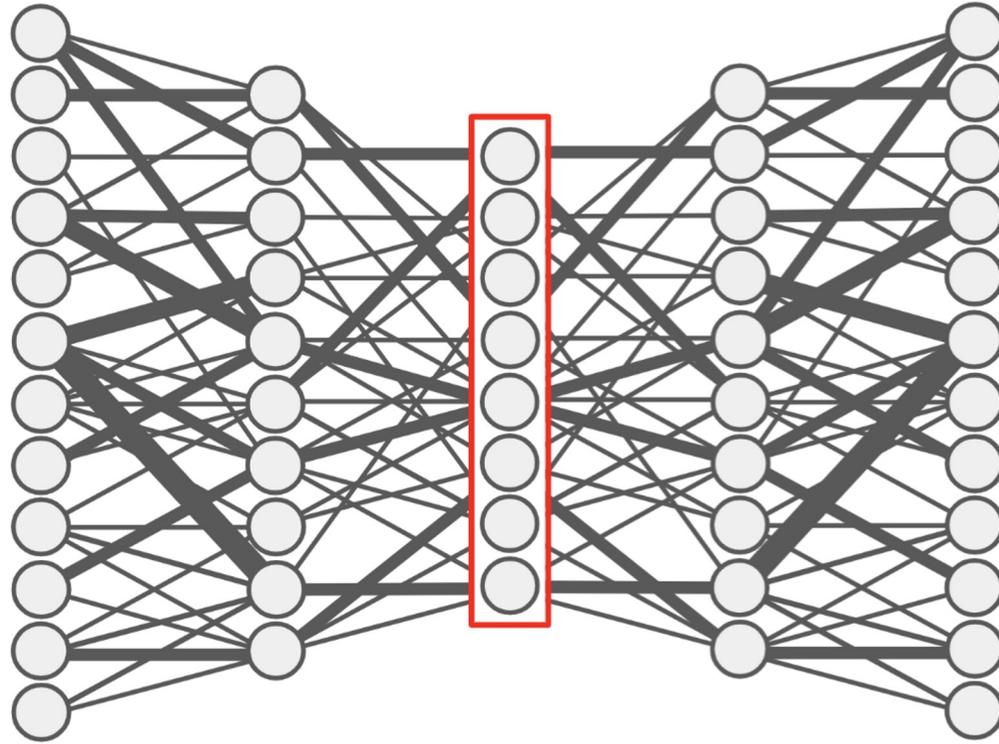
Generando imágenes



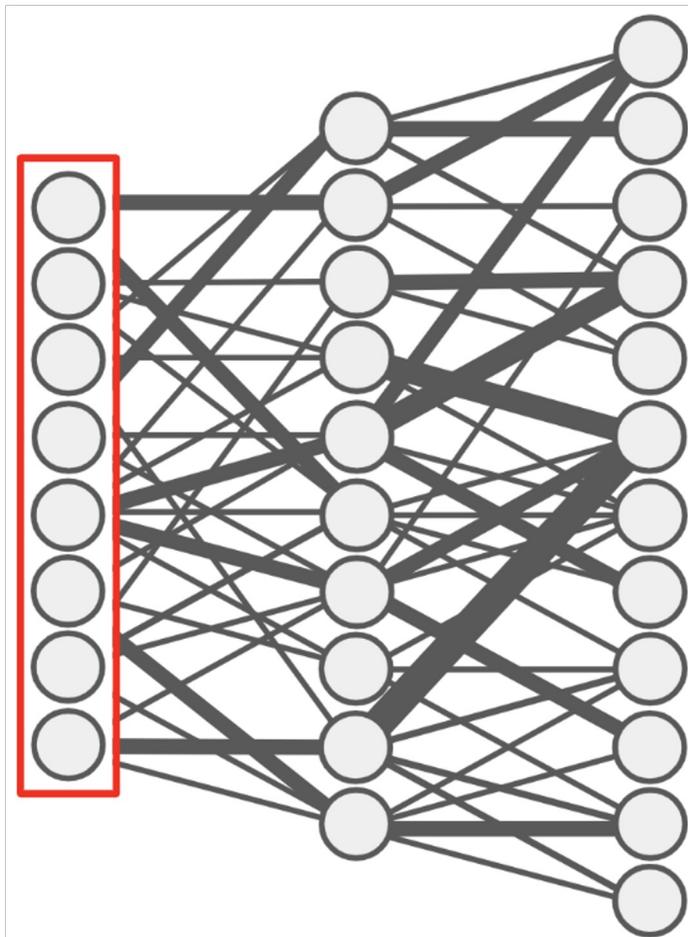


Entrenamos la nueva red  
repetiendo millones de veces...





Datos al azar





2014



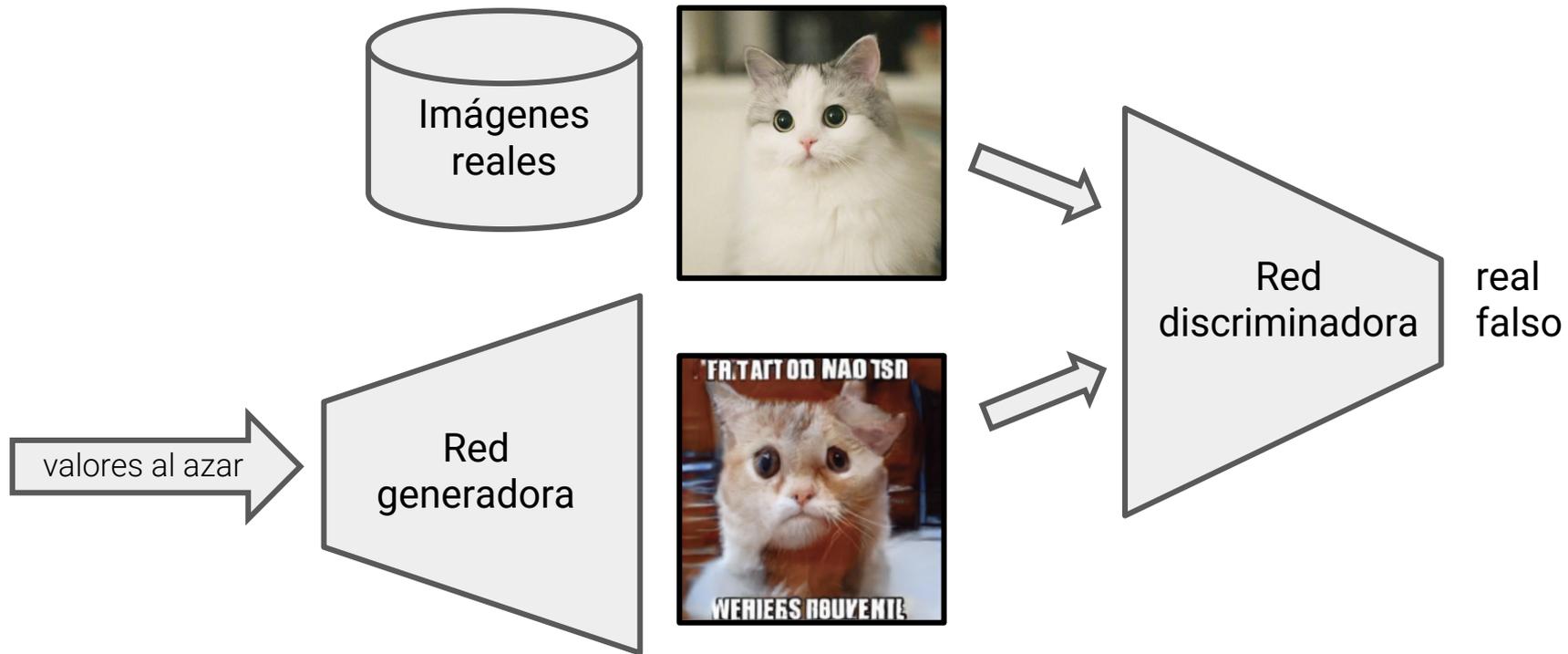
2015



2016



2017



# Living portraits



# Living portraits







# un sueño



# hecho realidad

# un sueño



hecho realidad



Select a face



Select a target

Keep fps

Keep audio

Keep frames

Many faces

Start

Destroy

Preview

# un sueño



# hecho



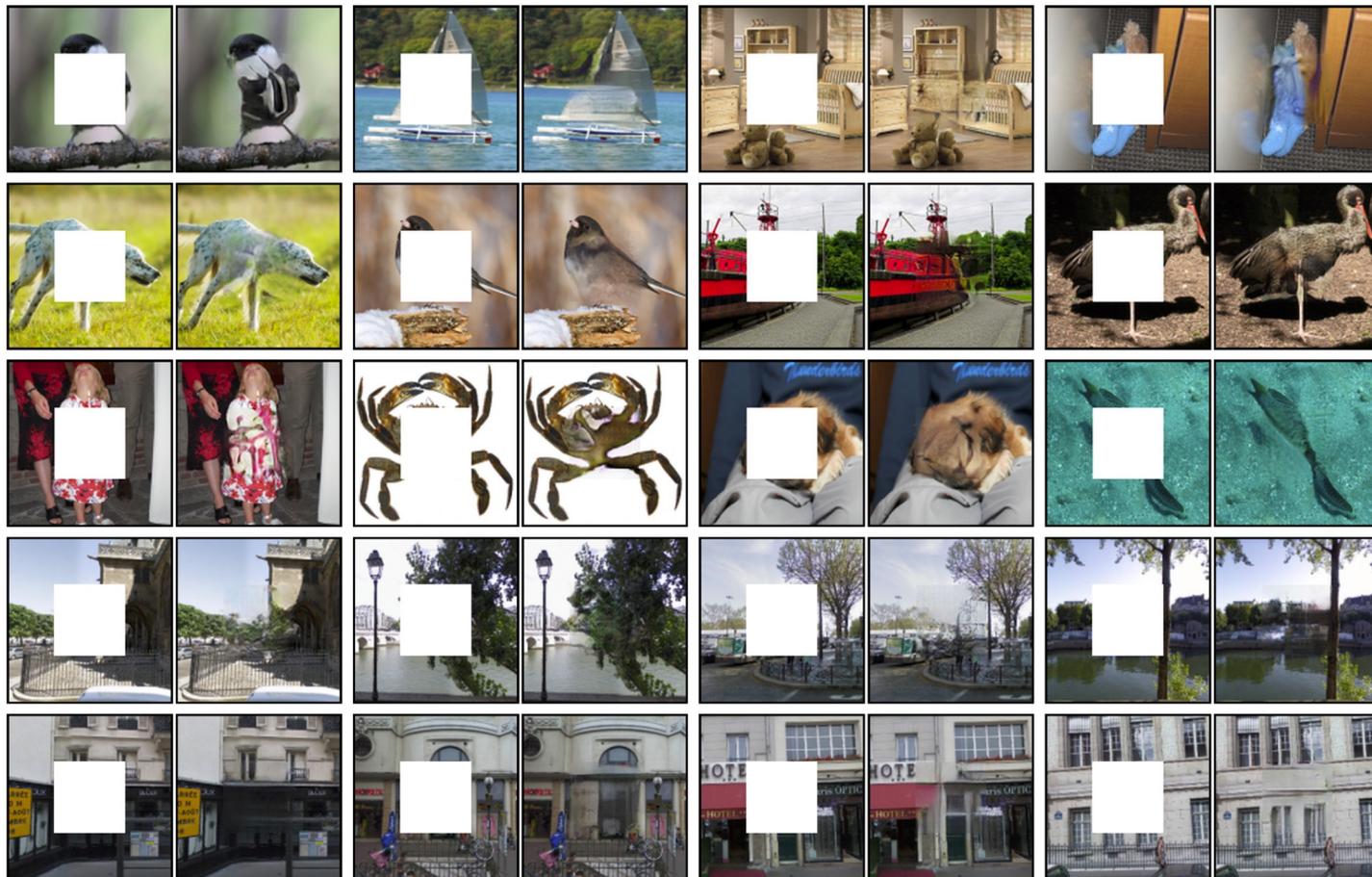




Image Normal Overlap





## A Generalist Agent

[Download](#)[View publication](#)[View blog](#)

### Abstract

Inspired by progress in large-scale language modeling, we apply a similar approach towards building a single generalist agent beyond the realm of text outputs. The agent, which we refer to as Gato, works as a multi-modal, multi-task, multi-embodiment generalist policy. The same network with the same weights can play Atari, caption images, chat, stack blocks with a real robot arm and much more, deciding based on its context whether to output text, joint torques, button presses, or other tokens. In this report we describe the model and the data, and document the current capabilities of Gato.

Authors

## Abstract

Inspired by progress in large-scale language modeling, we apply a similar approach towards building a single generalist agent beyond the realm of text outputs. The agent, which we refer to as Gato, works as a multi-modal, multi-task, multi-embodiment generalist policy. The same network with the same weights can play Atari, caption images, chat, stack blocks with a real robot arm and much more, deciding based on its context whether to output text, joint torques, button presses, or other tokens. In this report we describe the model and the data, and document the current capabilities of Gato.

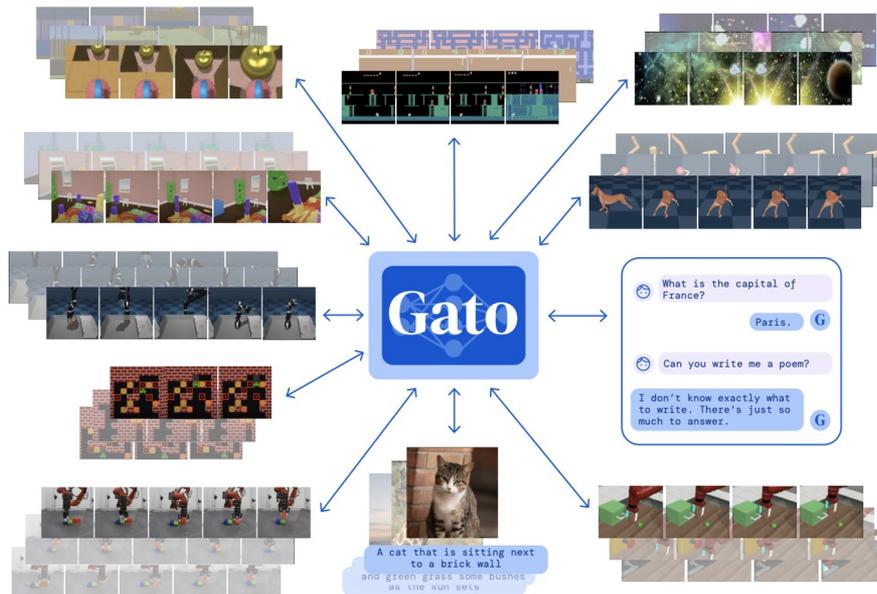


Figure 1: **A generalist agent.** Gato can sense and act with different embodiments across a wide range of environments using a single neural network with the same set of weights. Gato was trained on 604 distinct tasks with varying modalities, observations and action specifications.

---

# HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face

---

Yongliang Shen<sup>1,2\*</sup>, Kaitao Song<sup>2\*</sup>, Xu Tan<sup>2</sup>, Dongsheng Li<sup>2</sup>, Weiming Lu<sup>1</sup>, Yueting Zhuang<sup>1</sup>  
Zhejiang University<sup>1</sup>, Microsoft Research Asia<sup>2</sup>  
{syl, luwm, yzhuang}@zju.edu.cn, {kaitaosong, xuta, dongсли}@microsoft.com

## Abstract

Solving complicated AI tasks with different domains and modalities is a key step toward artificial general intelligence. While there are abundant AI models available for different domains and modalities, they cannot handle complicated AI tasks. Considering large language models (LLMs) have exhibited exceptional ability in language understanding, generation, interaction, and reasoning, we advocate that LLMs could act as a controller to manage existing AI models to solve complicated AI tasks and language could be a generic interface to empower this. Based on this philosophy, we present HuggingGPT, a framework that leverages LLMs (e.g., ChatGPT) to connect various AI models in machine learning communities (e.g., Hugging Face) to solve AI tasks. Specifically, we use ChatGPT to conduct task planning when receiving a user request, select models according to their function descriptions available in Hugging Face, execute each subtask with the selected AI model, and summarize the response according to the execution results. By leveraging the strong language capability of ChatGPT and abundant AI models in Hugging Face, HuggingGPT is able to cover numerous sophisticated AI tasks in different modalities and domains and achieve impressive results in language, vision, speech, and other challenging tasks, which paves a new way towards artificial general intelligence<sup>2</sup>.

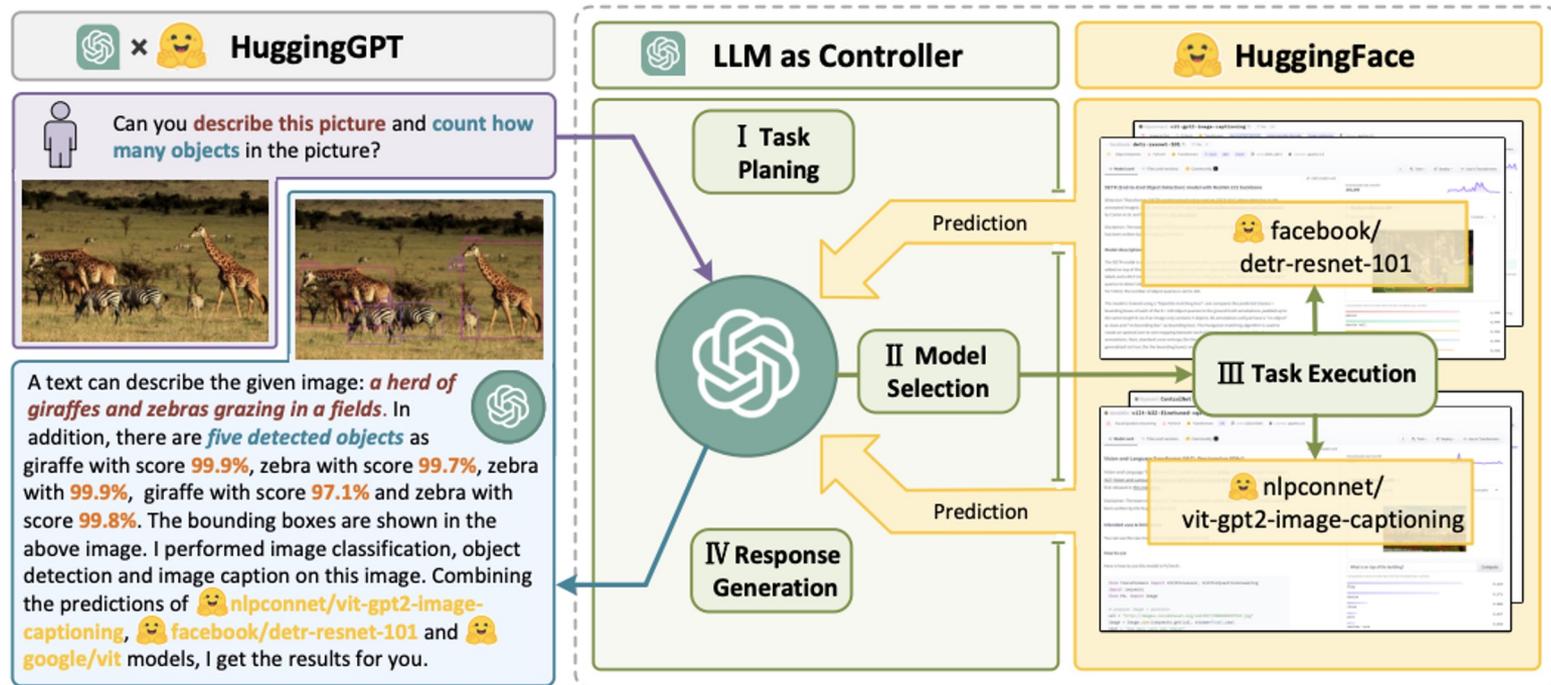
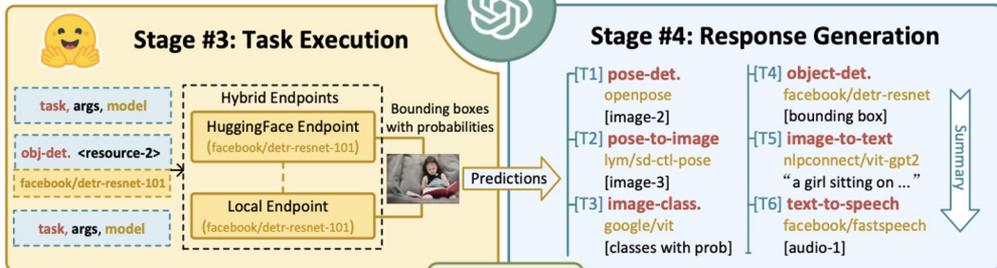
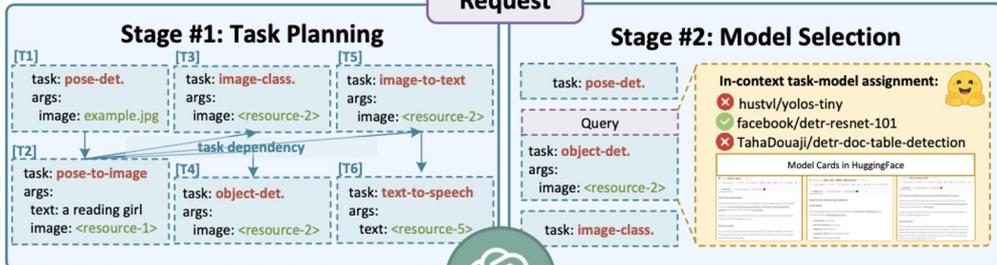


Figure 1: Language serves as an interface for LLMs (e.g., ChatGPT) to connect numerous AI models (e.g., those in Hugging Face) for solving complicated AI tasks. In this concept, an LLM acts as a controller, managing and organizing the cooperation of expert models. The LLM first plans a list of tasks based on the user request and then assigns expert models to each task. After the experts execute the tasks, the LLM collects the results and responds to the user.

Please generate an image where a girl is reading a book, and her pose is the same as the boy in the image example.jpg, then please describe the new image with your voice.

### Request



### Response






a girl sitting on a bed reading a book

[Text-1] = <resource-5>



[Image-1] = example.jpg [Image-2] = <resource-1> [Image-3] = <resource-2> [Image-4] [Audio-1]

The image you gave me is of "boy". The first thing I did was use 🤖 OpenCV's openpose model to analyze the pose of the boy in the image. I fed that image, specifically located at [Image-1], into the model to get an output image of the pose, located at [Image-2]. Then I take that pose and generate a new image, located at [Image-3], using 🤖 llyasviel/sd-controlnet-openpose and the instruction "a girl is reading". Next, I ran object detection on the image generated in the previous step, to get predictions of the objects present in the image, including the box location. These predictions were run through the 🤖 facebook/detr-resnet-101 model. The box was drawn on the image, which is located at [Image-4]. Then, I used the predictions of an image classification model 🤖 google/vit-base-patch16-224 and an image caption model 🤖 nlconnect/vit-gpt2-image-caption to generate the caption for newly generated image. It generated text: "a girl sitting on a bed reading a book". Lastly, I ran the 🤖 facebook/fastspeech2-en-es-speech model to convert the generated text into audio, located at [Audio-1].

La revolución digital está  
entrando una nueva fase,  
**de regreso de los bits a  
los átomos.**

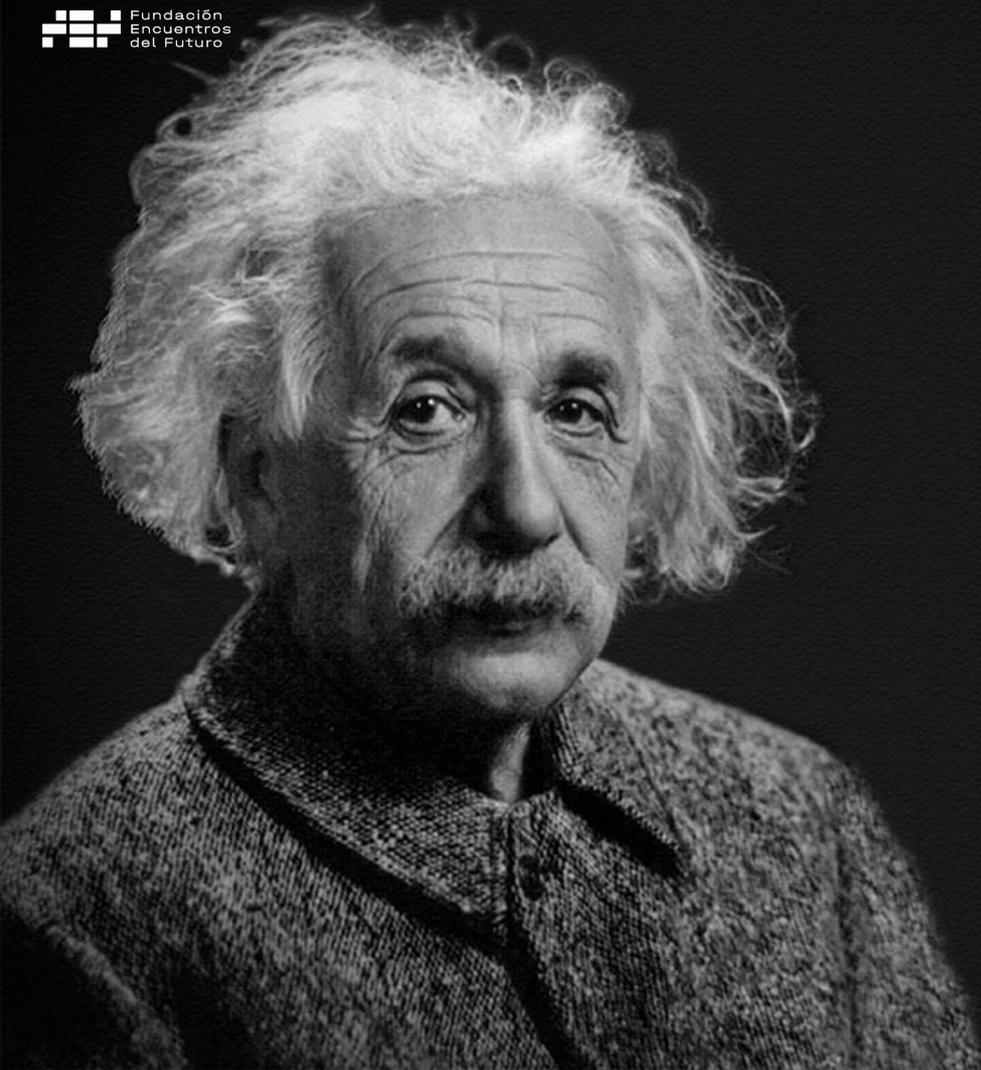
*From Atoms to Bits and Back: A Research  
Curation on Digital Technology and  
Agenda for Future Research*  
Bernd Schmitt, 2019



# átomos

# bits



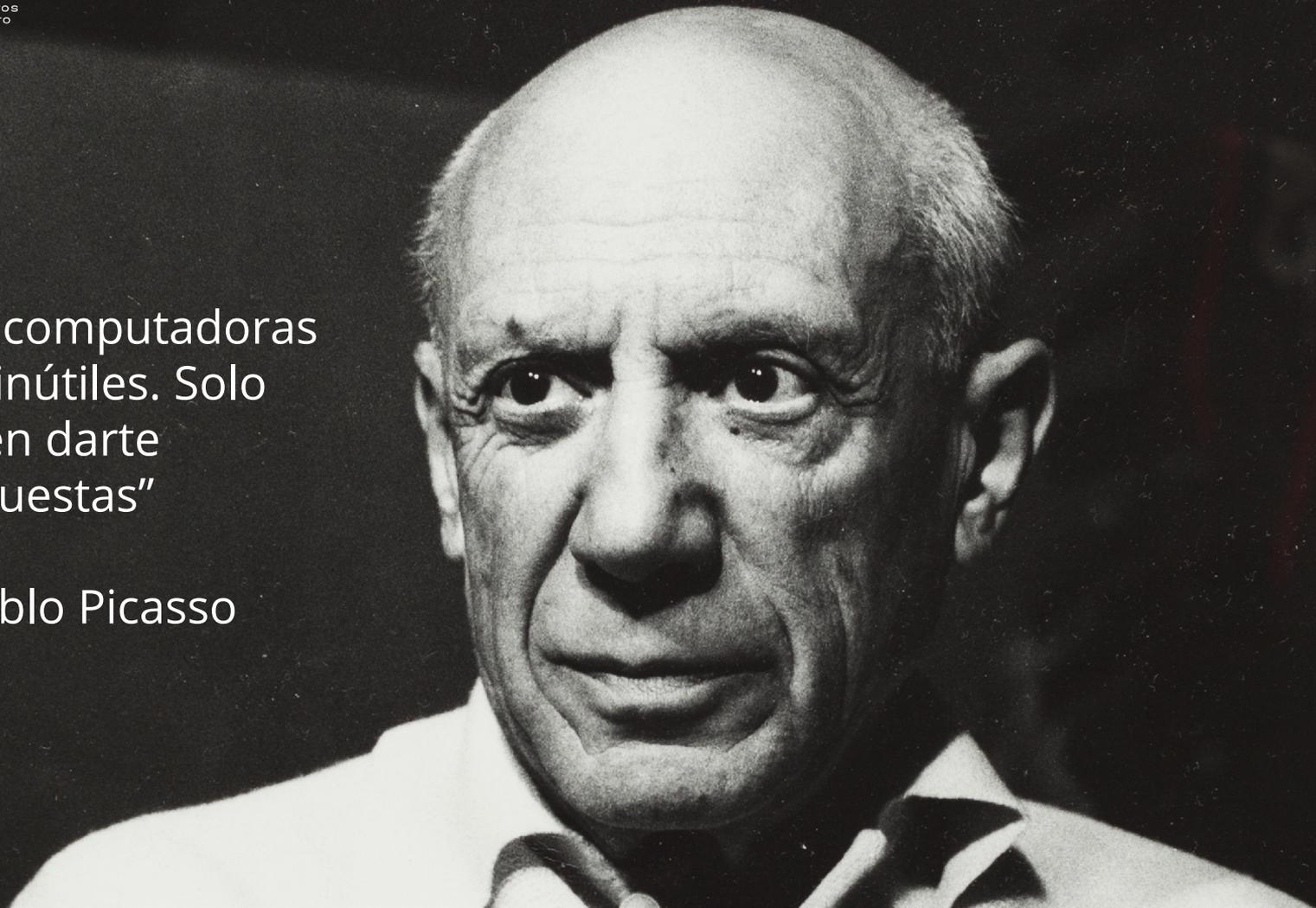


“El espíritu humano debe  
prevalecer sobre la tecnología”

-- Albert Einstein

“Las computadoras  
son inútiles. Solo  
saben darte  
respuestas”

-- Pablo Picasso



El **criterio**\_ es el  
arte de **\_elegir**

