

Saturdays.AI
Murcia



UCAM
UNIVERSIDAD
CATÓLICA DE MURCIA



INTRODUCCIÓN AL APRENDIZAJE POR REFUERZO

04 / 04 / 2020

Prof. Antonio Serrano Fernández

<https://antonioserranofz.com/>

ÍNDICE

1. INTRODUCCIÓN
2. TEORÍA Y CONCEPTOS BÁSICOS
3. APLICACIONES EN LA VIDA REAL
4. FRONTERA DEL CONOCIMIENTO
5. RECURSOS Y JUGADORES CLAVE

APRENDIZAJE SUPERVISADO

OBJETIVO: Crear una función capaz de predecir el valor de la variable objetivo

	PassengerId	Pclass	Sex	Age	SibSp	Parch	Fare	Cabin	Embarked	Survived
0	1	3	male	22.0	1	0	7.2500	NaN	S	0
1	2	1	female	38.0	1	0	71.2833	C85	C	1
2	3	3	female	26.0	0	0	7.9250	NaN	S	1
3	4	1	female	35.0	1	0	53.1000	C123	S	1
4	5	3	male	35.0	0	0	8.0500	NaN	S	0
5	6	3	male	NaN	0	0	8.4583	NaN	Q	0
6	7	1	male	54.0	0	0	51.8625	E46	S	0
7	8	3	male	2.0	3	1	21.0750	NaN	S	0
8	9	3	female	27.0	0	2	11.1333	NaN	S	1
9	10	2	female	14.0	1	0	30.0708	NaN	C	1
10	11	3	female	4.0	1	1	16.7000	G6	S	1
11	12	1	female	58.0	0	0	26.5500	C103	S	1
12	13	3	male	20.0	0	0	8.0500	NaN	S	0
13	14	3	male	39.0	1	5	31.2750	NaN	S	0
14	15	3	female	14.0	0	0	7.8542	NaN	S	0
15	16	2	female	55.0	0	0	16.0000	NaN	S	
16	17	3	male	2.0	4	1	29.1250	NaN	Q	
17	18	2	male	NaN	0	0	13.0000	NaN	S	
18	19	3	female	31.0	1	0	18.0000	NaN	S	
19	20	3	female	NaN	0	0	7.2250	NaN	C	



YES, WE CAN!

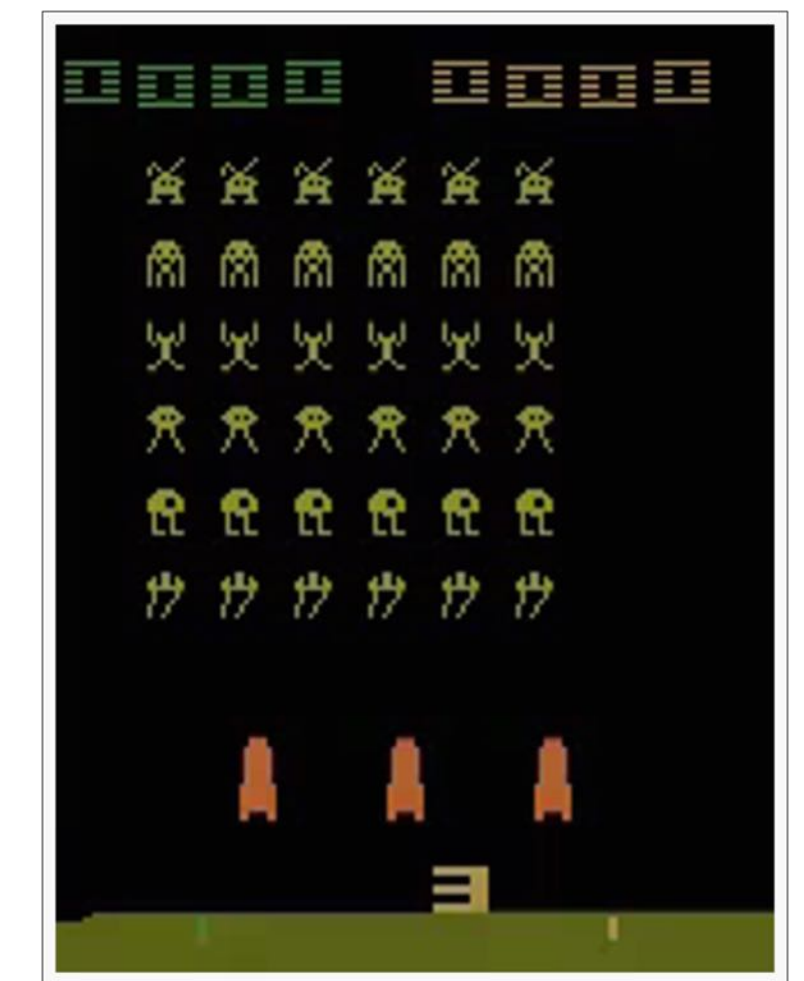
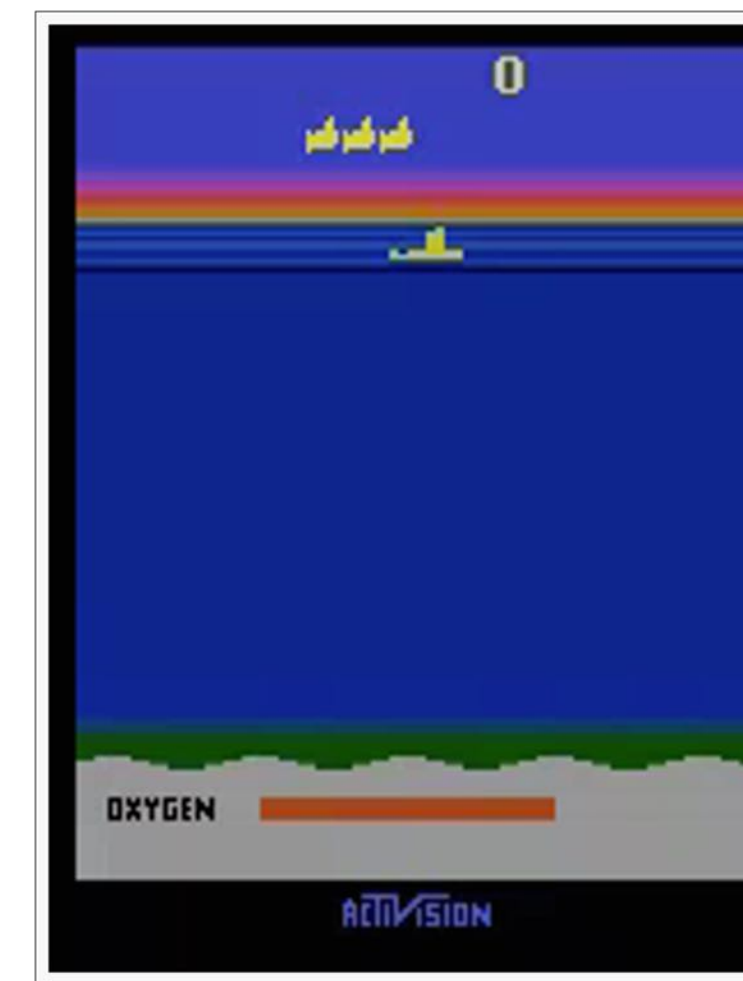


¿Y si pudiéramos utilizar esa función para enseñar a una máquina a tomar decisiones en un entorno y cumplir una tarea?

2013

DEEP Q-NETWORK

Las máquinas aprenden a jugar a los juegos de la consola Atari 2600

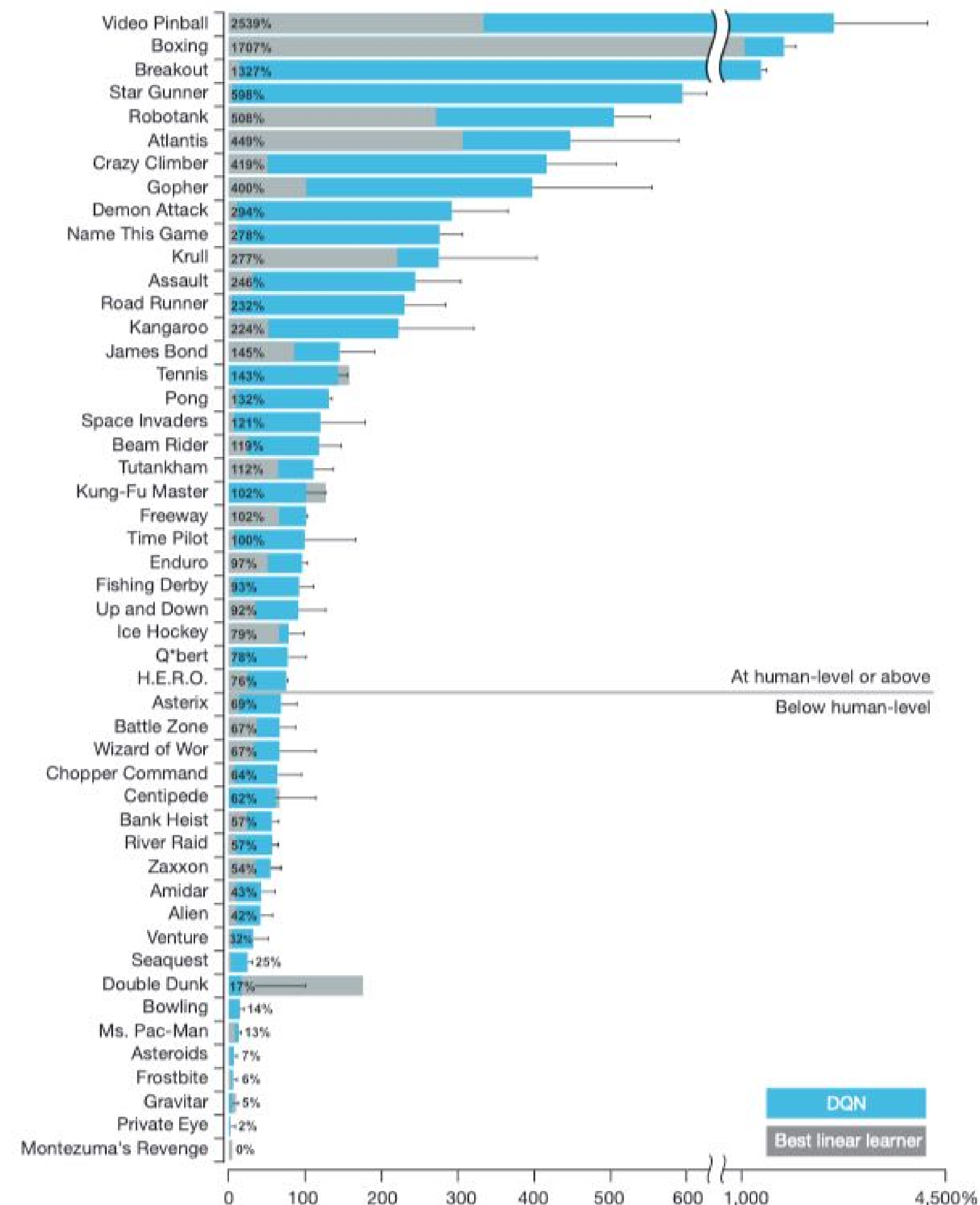


Mnih et al. (2015). Human-level control through deep reinforcement learning

2013

DEEP Q-NETWORK

Las máquinas aprenden a jugar a los juegos de la consola Atari 2600



2016

ALPHA GO

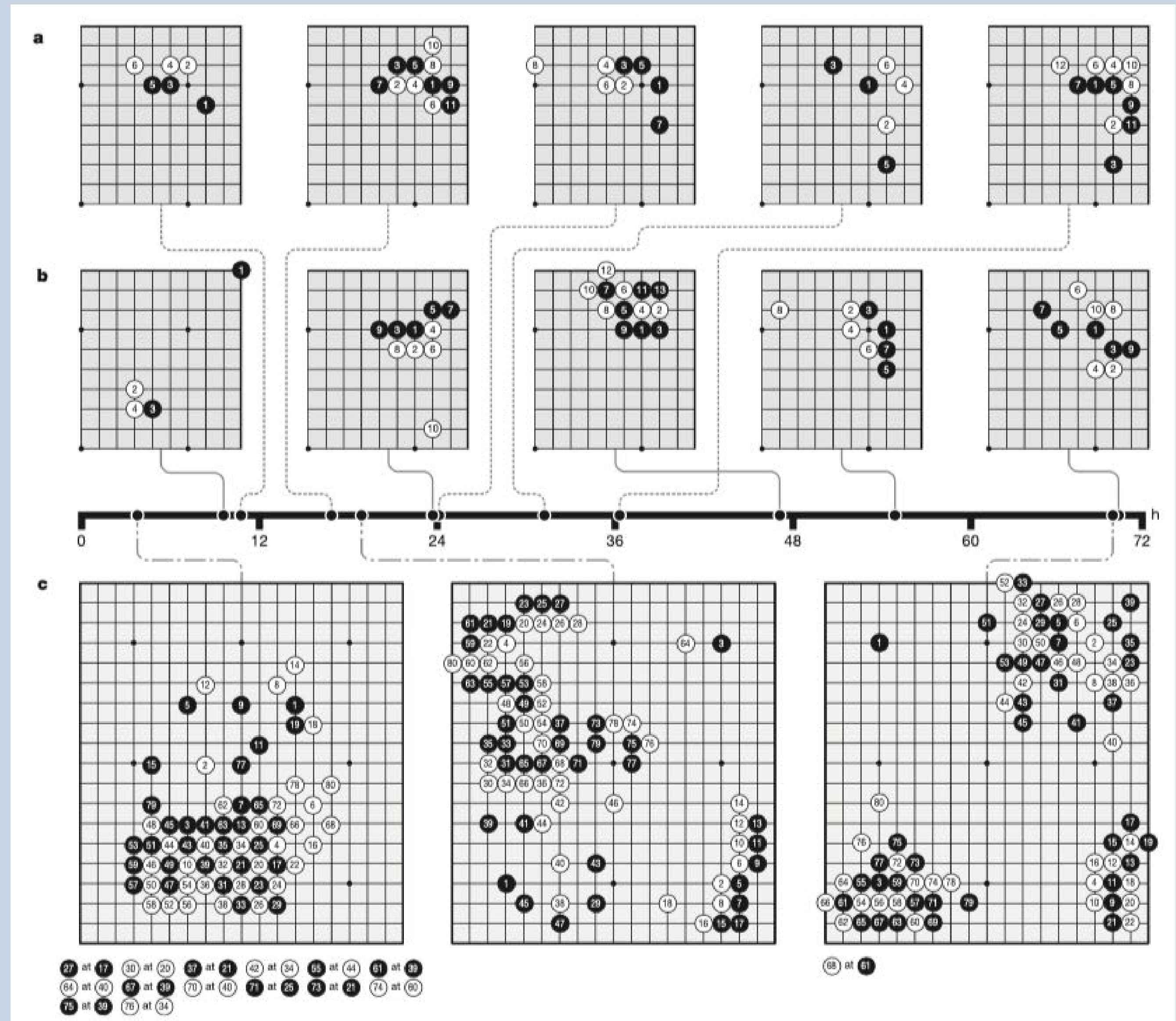
Un computador gana al campeón mundial de Go, Lee Sedol, por 4 a 1



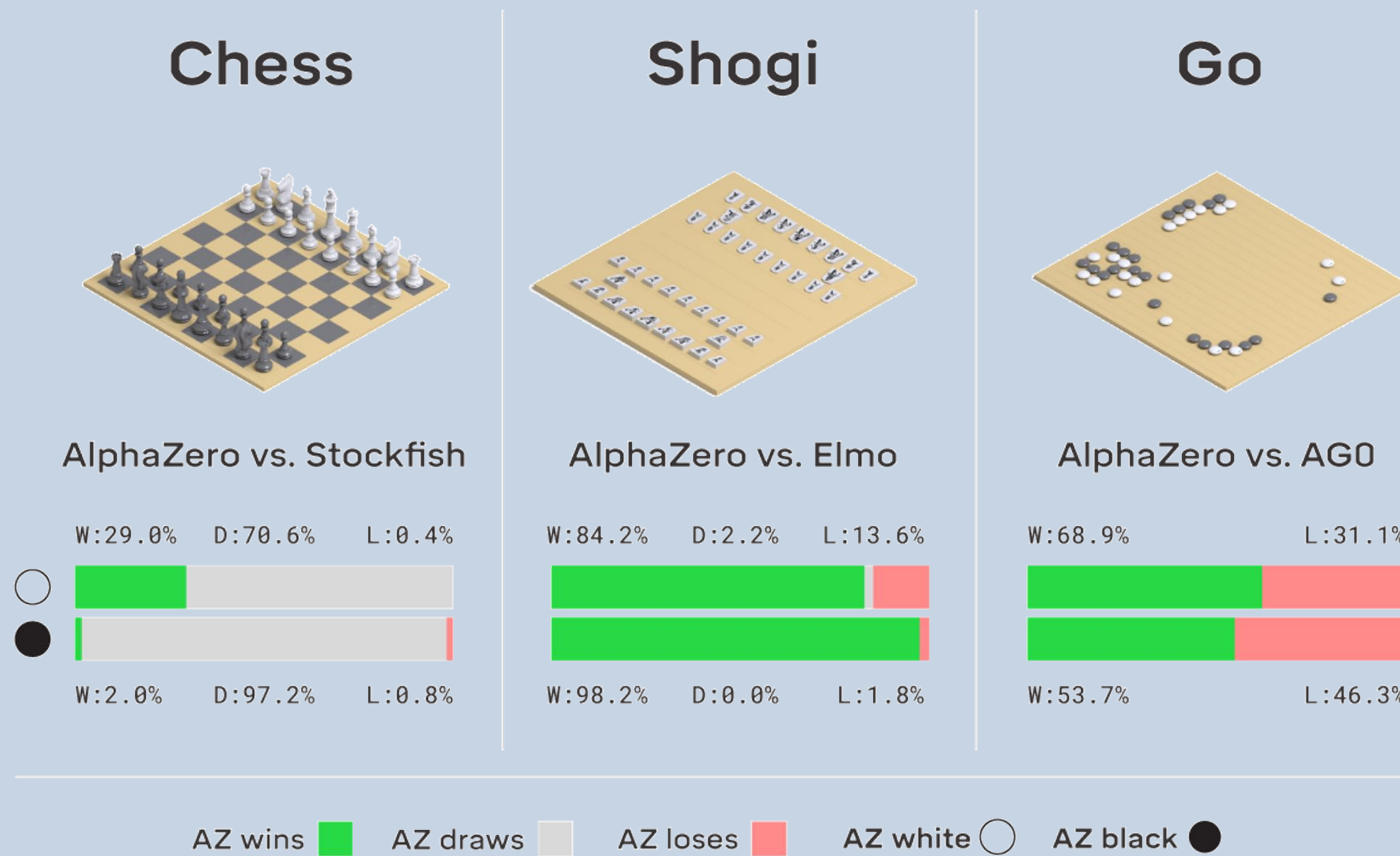
Silver et al. (2016). Mastering the game of Go with deep neural networks and tree search

2017 ALPHA ZERO

Versión mejorada de Alpha Go que
prescinde totalmente de datos de
partidas de jugadores humanos



Silver et al. (2017). Mastering the game of Go without human knowledge



2018 ALPHA ZERO

Versión mejorada de Alpha Go que prescinde totalmente de datos de partidas de jugadores humanos

Silver et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play

AlphaZero: Shedding new light on chess, shogi, and Go

2017-2019

DEEPSTACK, LIBRATUS & PLURIBUS

Las máquinas vencen a jugadores profesionales al póker
en la modalidad No-Limit Texas Hold'em



Imagen de [BlackRain79Poker](#)

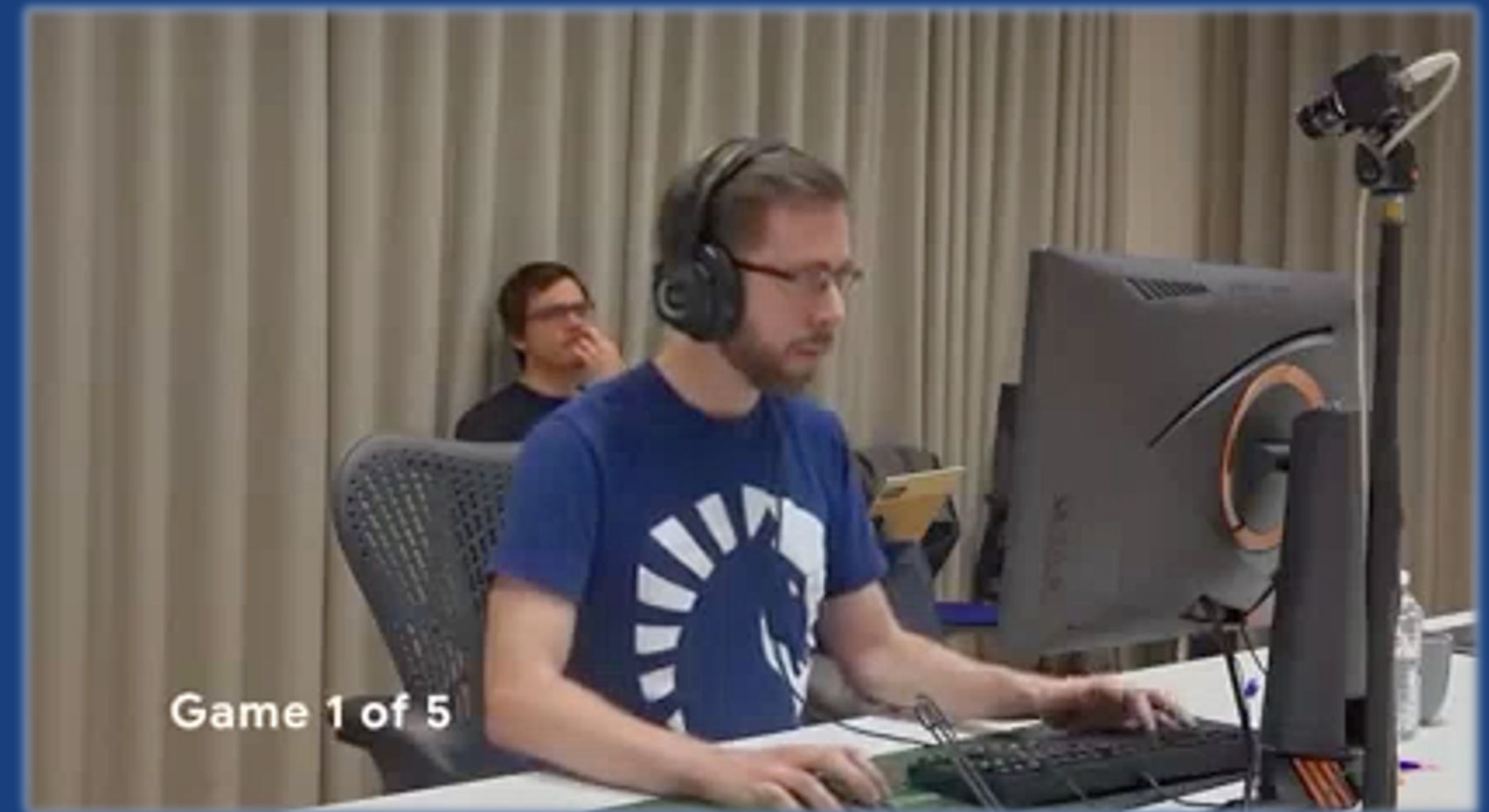
Moravčík et al. (2017). DeepStack: Expert-level artificial intelligence in heads-up no-limit poker

Brown & Sandholm (2018). Superhuman AI for heads-up no-limit poker: Libratus beats top professionals

Brown & Sandholm (2019). Superhuman AI for multiplayer poker

2019 ALPHA STAR

La máquina vence al equipo “Team Liquid” de Star Craft II por 10 a 1



Vinyals et al. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning

AlphaStar: Mastering the Real-Time Strategy Game StarCraft II

2019

AUTOMATIC DOMAIN RANDOMIZATION



Akkaya et al. (2019). Solving Rubik's Cube with a Robot Hand

Solving Rubik's Cube with a Robot Hand

El algoritmo ADR fue capaz de enseñar a una mano robótica a resolver el Cubo de Rubik de cero

ÍNDICE

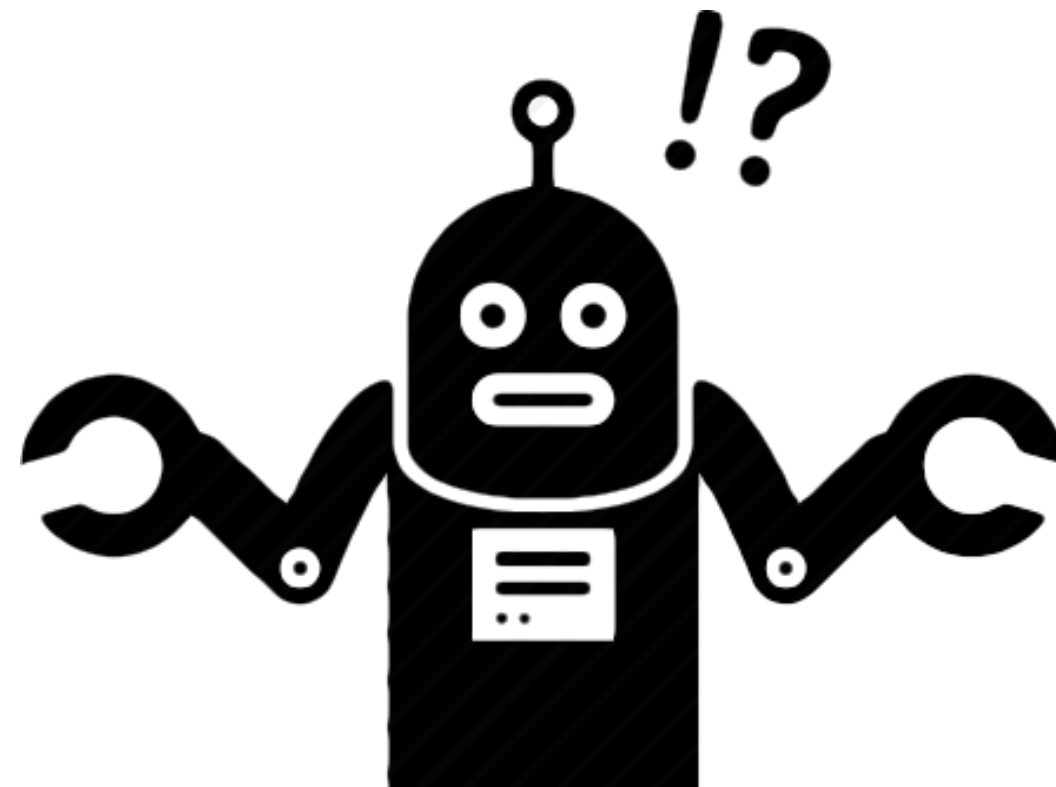
1. INTRODUCCIÓN
2. TEORÍA Y CONCEPTOS BÁSICOS
3. APLICACIONES EN LA VIDA REAL
4. FRONTERA DEL CONOCIMIENTO
5. RECURSOS Y JUGADORES CLAVE

DEEP REINFORCEMENT LEARNING

DEFINICIÓN

Respuesta rápida:

DEEP REINFORCEMENT LEARNING =
REINFORCEMENT LEARNING + DEEP LEARNING



REINFORCEMENT LEARNING

APRENDIZAJE POR REFUERZO

Disciplina dentro del Machine Learning que pretende enseñar a un **agente** a tomar acciones en un **entorno** para maximizar una **recompensa** acumulativa a largo plazo.

- ✎ Aprendizaje por prueba y error
- ✎ Recompensas ocasionales

75%

Deep RL aplica redes neuronales en ciertas partes de los algoritmos de RL:

- Función de política
- Función de valor
- Modelo de transición

DEEP LEARNING

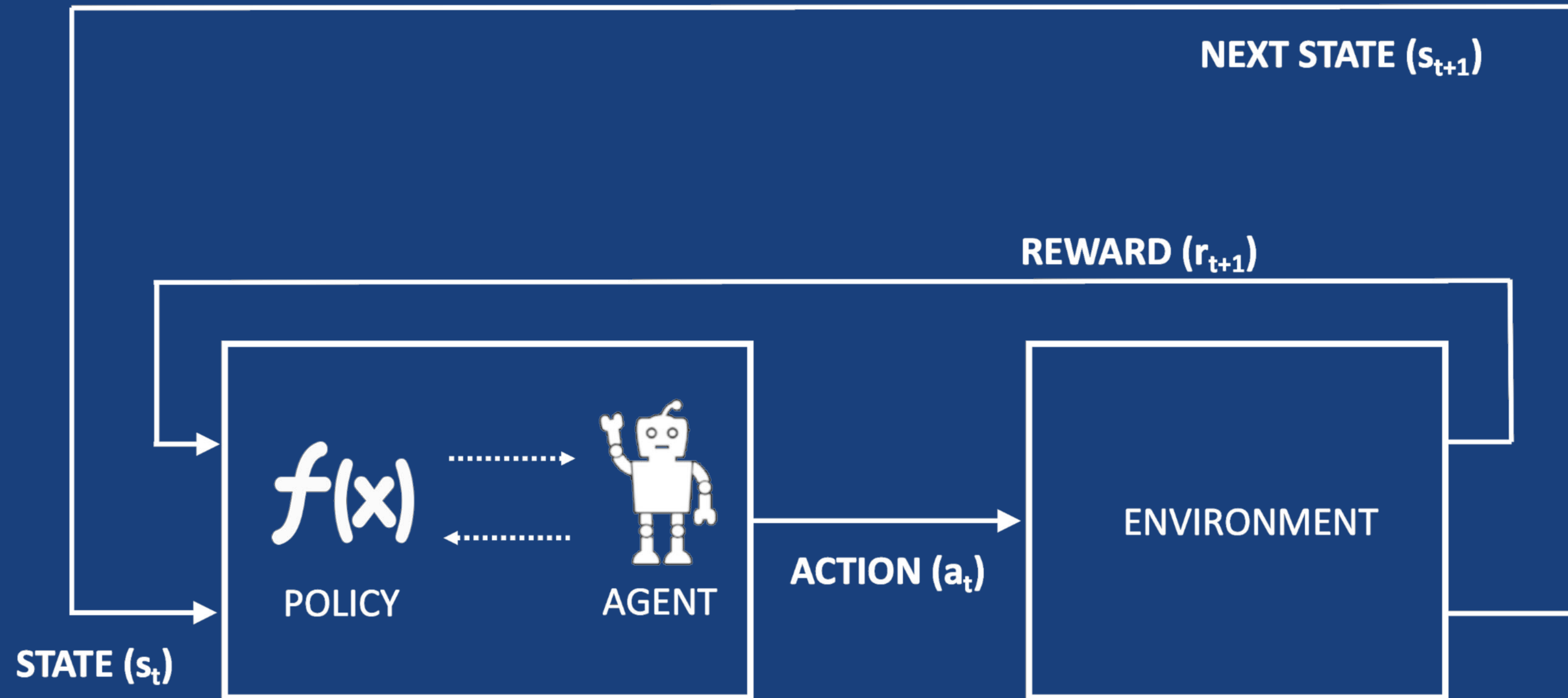
APRENDIZAJE PROFUNDO

Disciplina dentro del Machine Learning que utiliza métodos de **aprendizaje de representación** (sobre todo **redes neuronales**) para resolver ciertas tareas.

- ✎ Deep = redes neuronales con > 3 capas

25%

COMPONENTES BÁSICOS DE RL



- Nos centramos en tareas de horizonte finito y episódicas
- Dentro de cada episodio, el agente toma una acción por paso (time step)
- Como en todo flujo de trabajo de ML, existe un proceso de entrenamiento y otro de inferencia/predicción
- Se trata de encontrar la función de política óptima que maximiza la recompensa en el I/p

NOTA IMPORTANTE

- Dado el tiempo programado para la ponencia, no se cubrirá la parte más formal del marco teórico del RL, incluyendo:
 - Procesos de Decisión de Markov (MDPs)
 - Definición de la función de política
 - El descuento γ en las recompensas
 - Exploración vs. explotación
 - Algoritmos basados en modelos vs. libres
 - Algoritmos basados en la función de política vs. función de valor
 - La iteración de política generalizada (GPI):
 - Definición de la función de valor
 - Ecuación de Bellman
 - Algoritmos básicos de RL:
 - Bandido multibrazos
 - Monte Carlo
 - Programación Dinámica
 - Diferencia Temporal (TD)
 - Algoritmos avanzados:
 - DQN
 - DDPG
 - Policy Gradient
 - PPO
 - A3C
 - Etc.
- Para profundizar en estos contenidos (salvo algoritmos avanzados) → Ver manual de Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*, Capítulos 1-6 y 9

DIFERENCIAS ENTRE APRENDIZAJE SUPERVISADO Y RL

- NOCIÓN DE AGENTE Y ENTORNO
- TOMA DE DECISIONES SECUENCIALES
- NO HAY ETIQUETAS, SINO RECOMPENSAS
- RECOMPENSA CON RETRASO
- MUESTRAS GENERADAS POR INTERACCIÓN CON EL ENTORNO SOBRE LA MARCHA
- BÚSQUEDA DE LA SOLUCIÓN MEDIANTE PRUEBA Y ERROR
- ÁMBITO DE APLICACIÓN MÁS CEÑIDO A LA IA

DIFERENCIAS ENTRE APRENDIZAJE SUPERVISADO Y RL

Fridman (2019). Introduction to Deep Reinforcement Learning



COMPLEJIDAD DEL RL

Seamos sinceros... RL es una disciplina compleja y difícil porque es muy general y abarca problemas muy diferentes

- FOMDP vs. POMDP
- Dinámicas del entorno estacionarias vs. no estacionarias
- Tareas de horizonte finito, indefinido o infinito
- Tareas episódicas vs. no episódicas
- Estados y/o acciones discretos vs. continuos
- Soluciones tabulares vs. aproximadas
- Algoritmos basados en modelos vs. libres
- ...

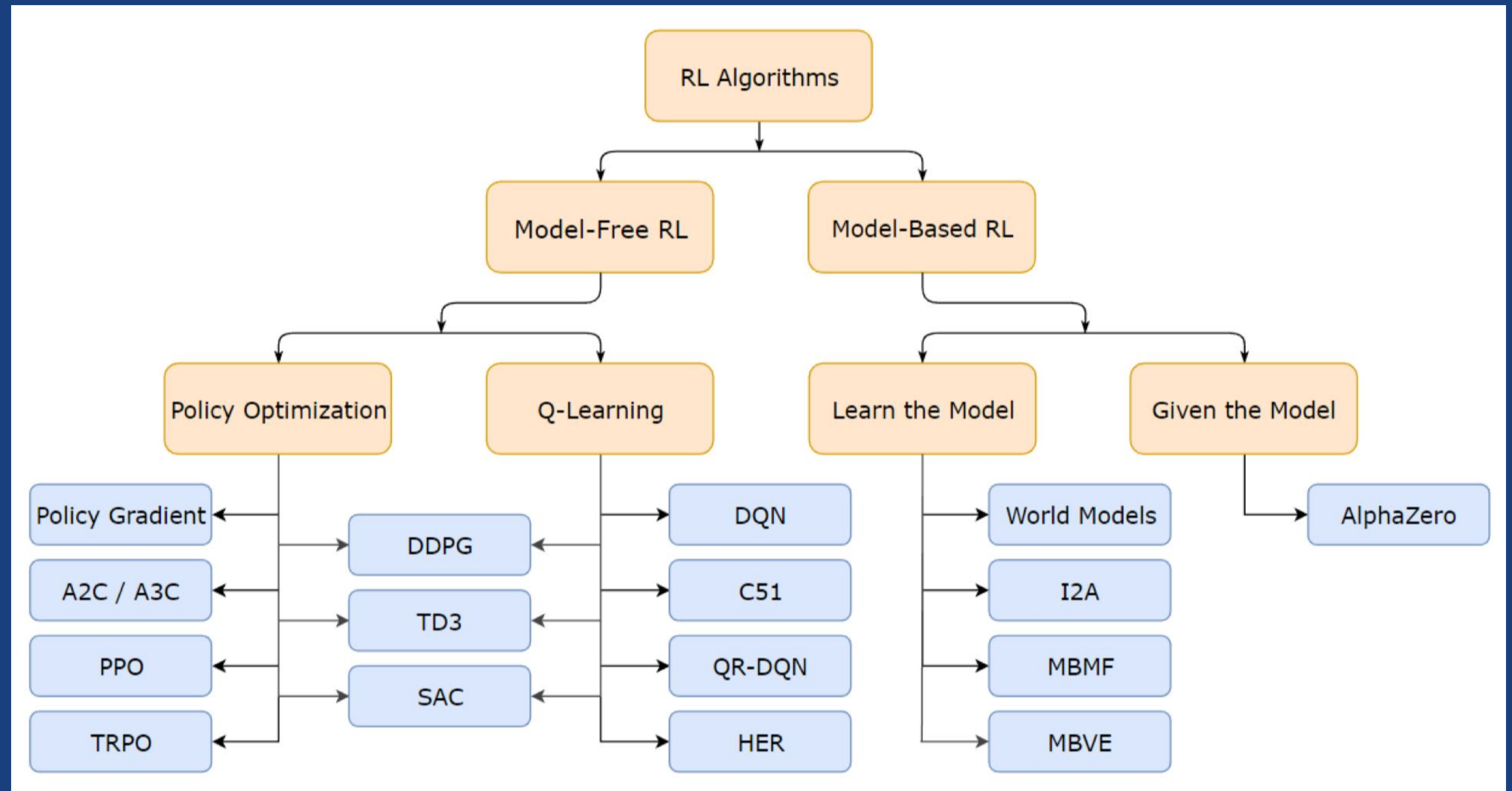
COMPLEJIDAD DEL RL

Seamos sinceros... RL es una disciplina compleja y difícil porque es muy general y abarca problemas muy diferentes

- Algoritmos basados en funciones de valor o de política
- Algoritmos off-policy vs. on-policy
- Políticas deterministas vs. estocásticas
- Políticas según equilibrio exploración/explotación: greedy determinista, ϵ -soft, ϵ -greedy...
- Grado de bootstrapping o TD en algoritmos de funciones de valor
- Entornos con una sola tarea vs. multitarea
- Entornos con un solo agente vs. multiagente
- Etc.

COMPLEJIDAD DEL RL

Ello da lugar a
numerosos
algoritmos



[Fridman \(2019\). Introduction to Deep Reinforcement Learning](#)

COMPLEJIDAD DEL RL



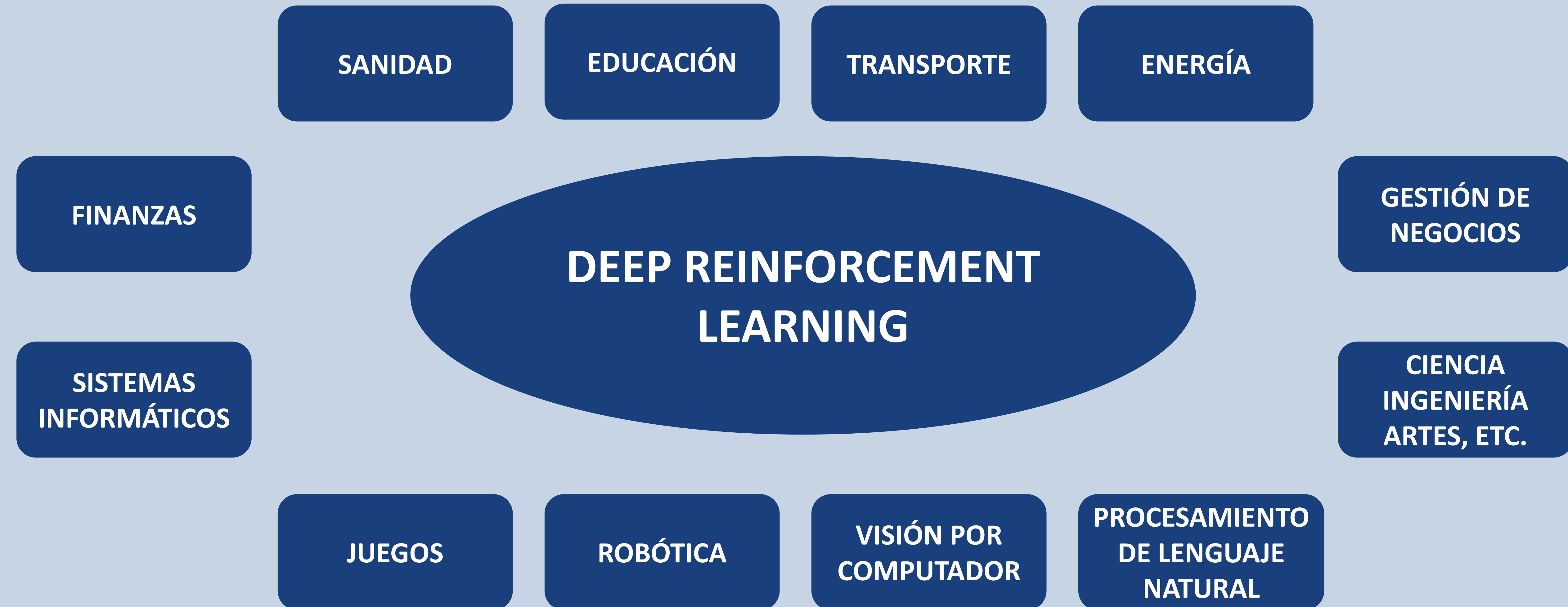
BostonDynamics (2017).
What's new, Atlas?

PERO A LA PAR ES MUY
GRATIFICANTE!!!

ÍNDICE

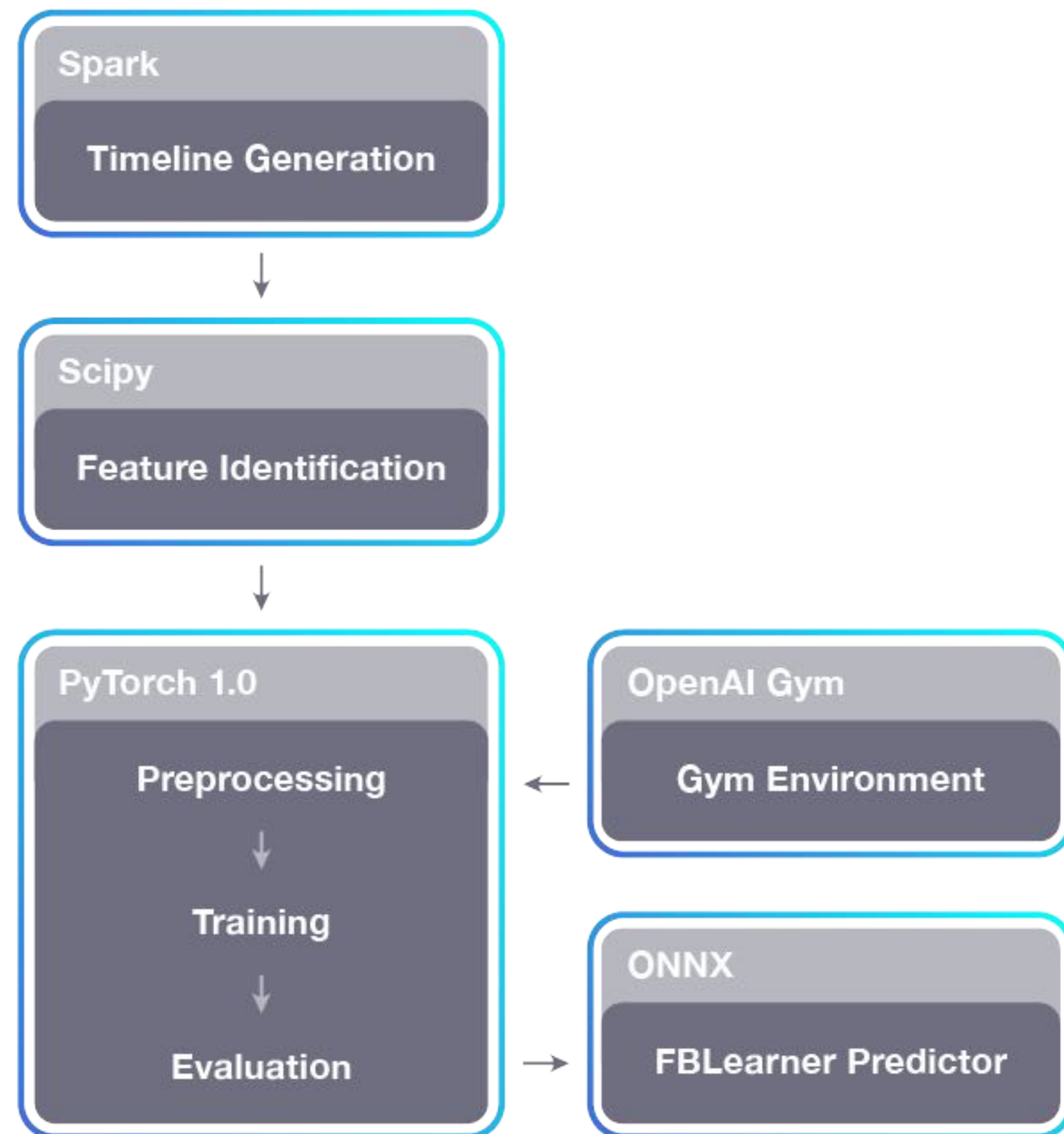
1. INTRODUCCIÓN
2. TEORÍA Y CONCEPTOS BÁSICOS
3. APLICACIONES EN LA VIDA REAL
4. FRONTERA DEL CONOCIMIENTO
5. RECURSOS Y JUGADORES CLAVE

ÁMBITOS DE APLICACIÓN DEL DEEP RL



SISTEMAS DE PRODUCCIÓN

FACEBOOK HORIZON



- Plataforma de código abierto de extremo a extremo que utiliza RL para optimizar sistemas en entornos de producción a gran escala
- Incluye frameworks de PyTorch, Caffe2 y Spark
- Aplicaciones probadas en Facebook:
 - Personalizar M sugerencias de Messenger
 - Proveer notificaciones más útiles
 - Optimizar video en streaming

Gauci et al. (2019). Horizon: Facebook's Open Source Applied Reinforcement Learning Platform

CONDUCCIÓN AUTÓNOMA

WAYVE



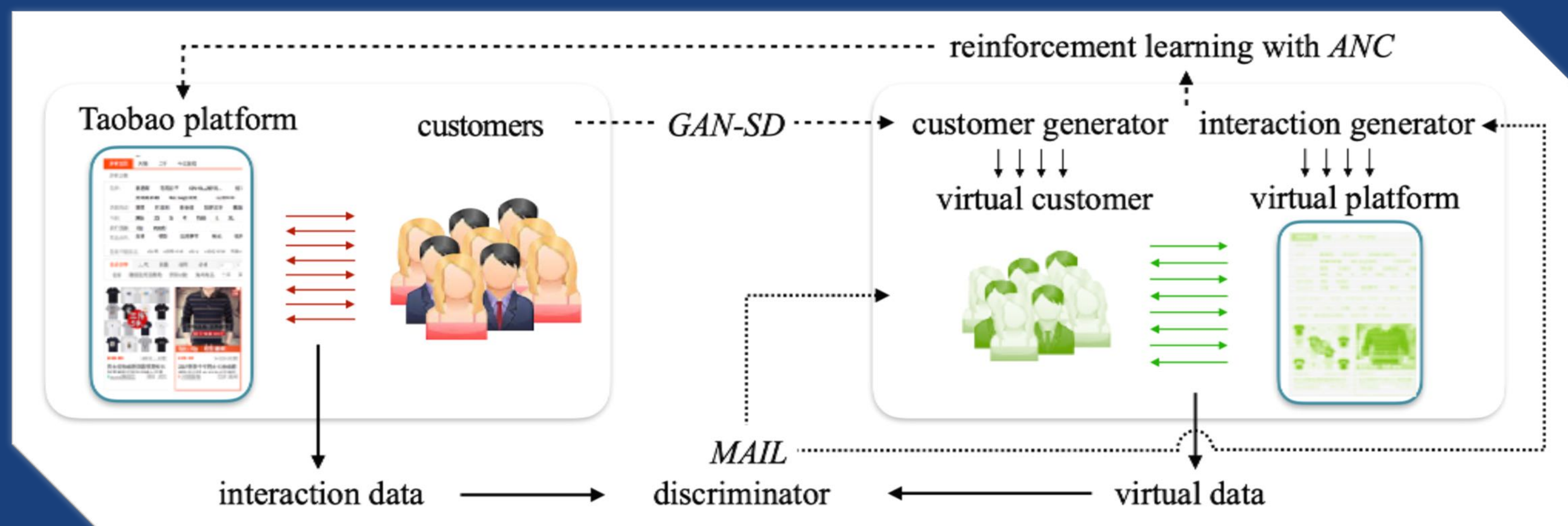
- Empresa de Cambridge (UK) dedicada a la conducción autónoma
- No sólo aplican conocimientos de Visión por computador y DL, sino también de Deep RL
- En el video se identifica claramente la primera y la última iteración y se ve cómo el automóvil aprende a torcer el volante para no desviarse del carril

[Wayve \(2018\). Learning to drive in a day](#)

SECTOR MINORISTA

ALIBABA

- Aplicación de Deep RL para una mejor búsqueda de productos en Taobao con el fin de incrementar las ventas
- Se hizo una una réplica de la tienda electrónica a pequeña escala como entorno de RL
- Combinación de modelos GAN-SD y MAIL
- Se crearon políticas que alcanzaban un 2% más de rentabilidad en ventas



[Shi et al. \(2018\). Virtual-Taobao: Virtualizing Real-world Online Retail Environment for Reinforcement Learning](#)

ROBÓTICA OSARO



[Osaro \(2017\). Industrial robotics and deep reinforcement learning](#)

- Empresa de San Francisco especializada en software de IA para automatización industrial, sobre todo Deep RL
- En el video, se entrenan a brazos robóticos a manipular objetos mediante algoritmos de Imitation Learning y Deep RL

EXTRACCIÓN PETROLÍFERA

SHELL

- Algoritmos de Deep RL para guiar las perforaciones por el subsuelo en busca de crudo
- Entrenados con datos históricos de perforación de Shell y exploración simulada
- El operario es capaz de comprender el entorno del subsuelo con mayor precisión, lo acelera el proceso de perforación y genera un menor desgaste, roturas y daños a la maquinaria

Operario instalando una guía de broca en una plataforma petrolífera en Texas
Imagen de [Transport Topics](#)



DESCUBRIMIENTO DE FÁRMACOS

UCAM



Antonio Serrano



Dr. Baldomero Imbernón



Dr. Horacio Pérez-Sánchez



UCAM
UNIVERSIDAD
CATÓLICA DE MURCIA



Dr. José M. Cecilia



Dr. Andrés Bueno-Crespo



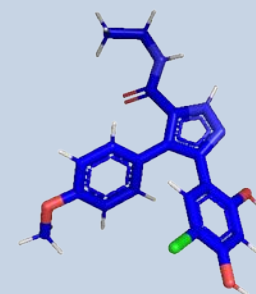
Dr. José L. Abellán

[Serrano et al. \(2018\). Accelerating Drugs Discovery with Deep Reinforcement Learning: An Early Approach](#)

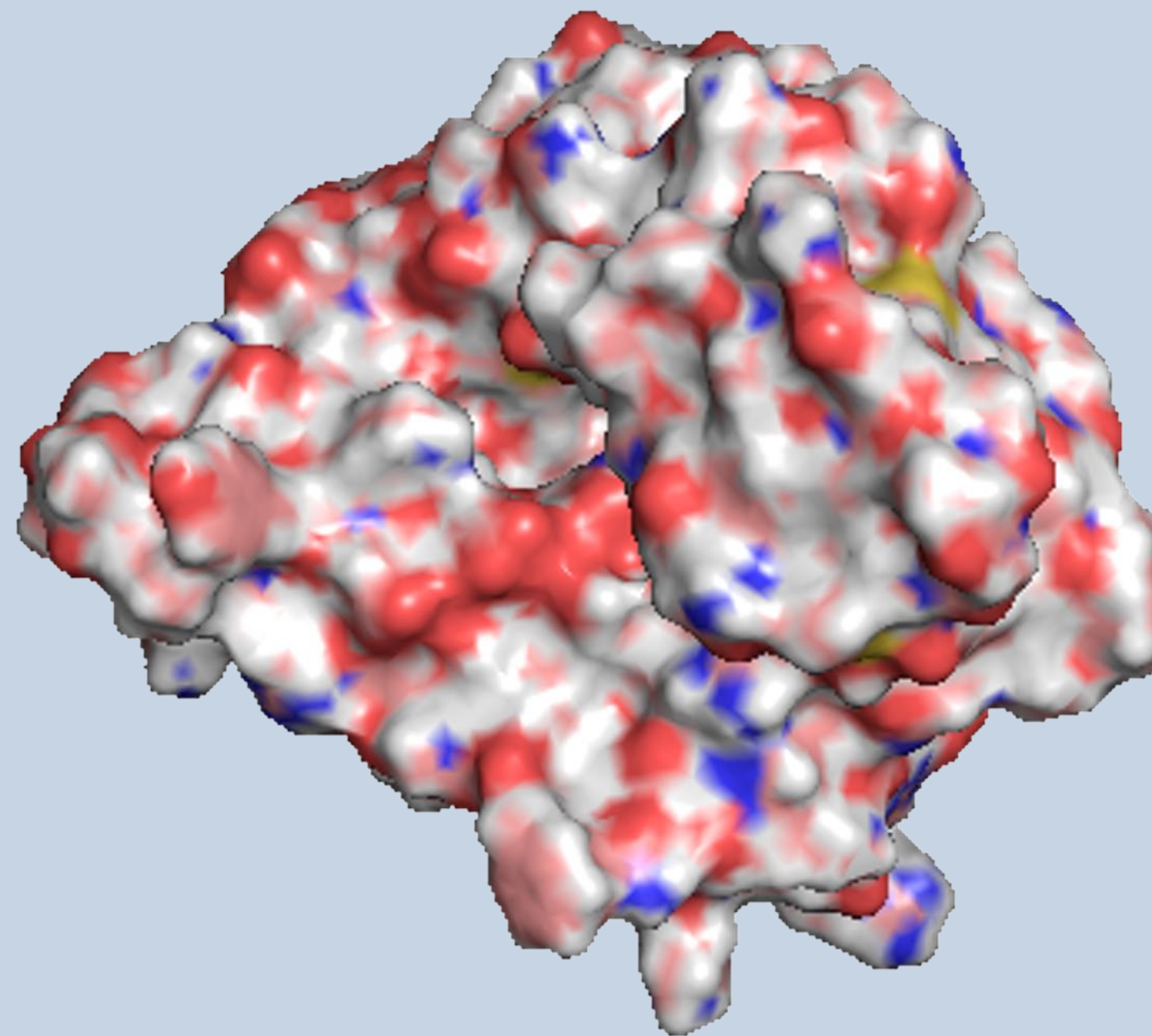
DESCUBRIMIENTO DE FÁRMACOS

UCAM

Técnica de Docking: buscar en un entorno virtual el punto de anclaje entre la molécula candidata a medicamento y la molécula receptora.



LIGANDO O CANDIDATO
FARMACOLÓGICO



MOLÉCULA RECEPTORA
(PROTEÍNA)

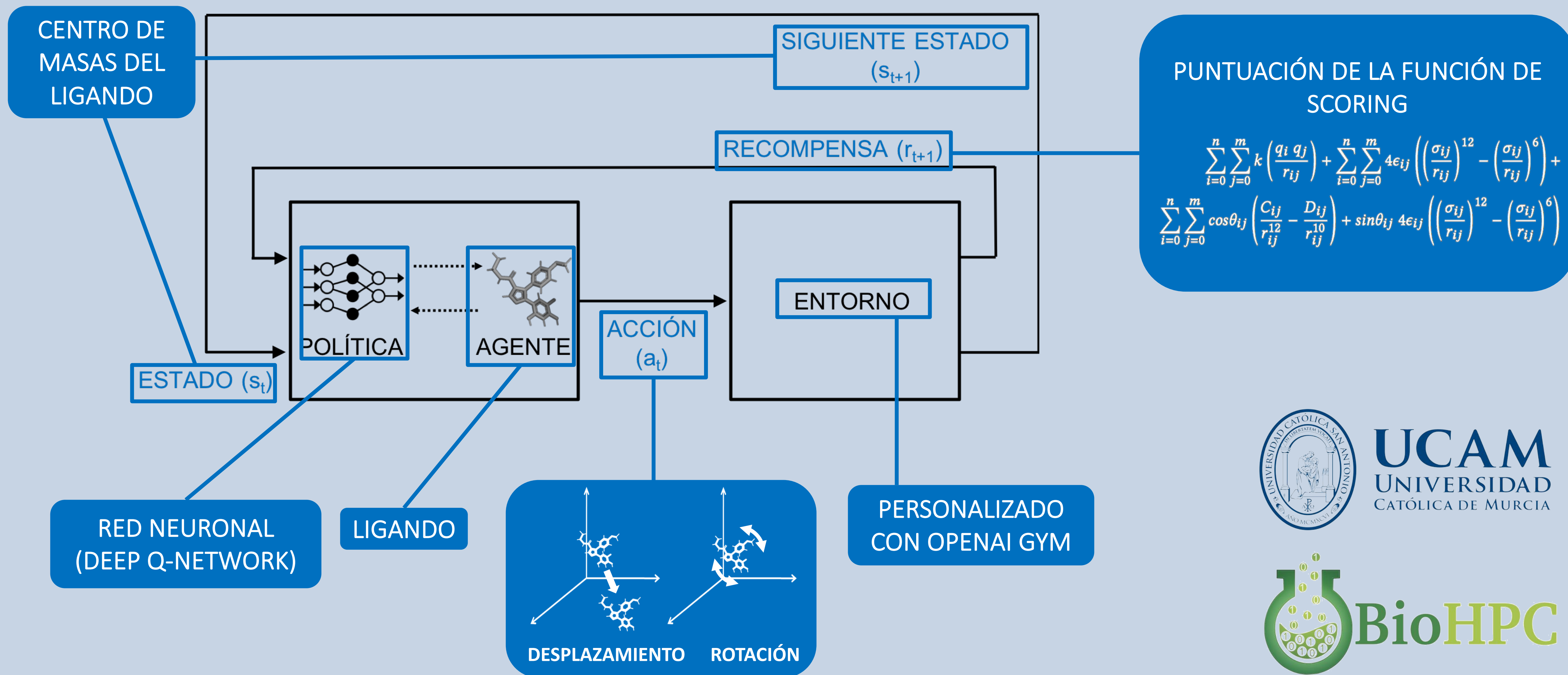


UCAM
UNIVERSIDAD
CATÓLICA DE MURCIA



DESCUBRIMIENTO DE FÁRMACOS

UCAM



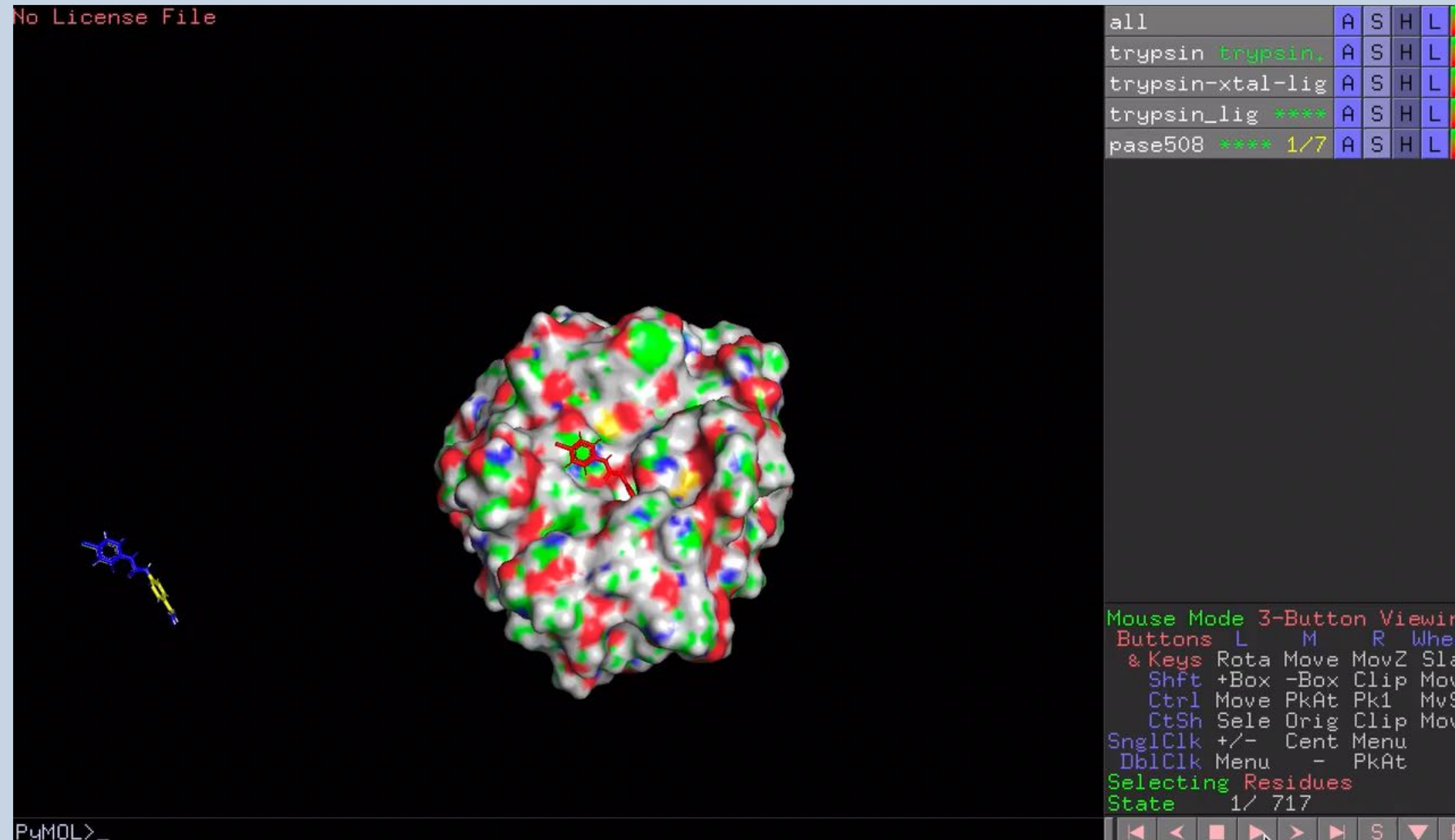
UCAM
UNIVERSIDAD
CATÓLICA DE MURCIA



DESCUBRIMIENTO DE FÁRMACOS

UCAM

VIDEO DEL AGENTE DURANTE EL ENTRENAMIENTO



UCAM
UNIVERSIDAD
CATÓLICA DE MURCIA



ÍNDICE

1. INTRODUCCIÓN
2. TEORÍA Y CONCEPTOS BÁSICOS
3. APLICACIONES EN LA VIDA REAL
4. FRONTERA DEL CONOCIMIENTO
5. RECURSOS Y JUGADORES CLAVE

ÁREAS DE INVESTIGACIÓN EN DEEP RL EN LA ACTUALIDAD

[Arulkumaran \(2017\)](#)

[Lange \(2020\)](#)



ÍNDICE

1. INTRODUCCIÓN
2. TEORÍA Y CONCEPTOS BÁSICOS
3. APLICACIONES EN LA VIDA REAL
4. FRONTERA DEL CONOCIMIENTO
5. RECURSOS Y JUGADORES CLAVE

RECURSOS

HABILIDADES REQUERIDAS EN DEEP RL

1. EXPERIENCIA CON **PYTHON** INTERMEDIA O AVANZADA

- Programación orientada a objetos
- Capaz de escribir bucles anidados
- Capaz de leer y comprender código de terceros

2. CONOCIMIENTO BÁSICO DE **ESTADÍSTICA Y ÁLGEBRA LINEAL**

- Probabilidades y distribuciones
- Multiplicación matricial y derivadas

RECURSOS

HABILIDADES REQUERIDAS EN DEEP RL

3. CONOCIMIENTO INTERMEDIO DE TÉCNICAS DE DEEP LEARNING

- Capaz de describir el algoritmo de backpropagation
- Haber visto/trabajado con arquitecturas de redes neuronales (como MLPs y CNNs)

4. CONOCIMIENTO DE FRAMEWORKS DE DEEP LEARNING

- Haber visto/trabajado con frameworks de DL como Tensorflow, Keras o Pytorch

5. CONOCIMIENTO AVANZADO DE INGLÉS (EL MÁS IMPORTANTE!!!)

CURSOS

- Curso de RL David Silver (DeepMind) **[gratuito]**
- Asignatura CS 285 sobre Deep RL de Sergey Levine (UC Berkeley) **[gratuito]**
- Nanodegree de Deep RL en Udacity **[4 meses x 180 euros/mes]**
- Especialización en DL de Andrew Ng (Stanford) **[172 euros]**

MANUALES

- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press **[gratuito, a falta de las soluciones a los ejercicios]**
- Morales, M. (2020). *Grokking Deep Reinforcement Learning*. Manning **[40 dólares]**



CONGRESOS TOP-TIER

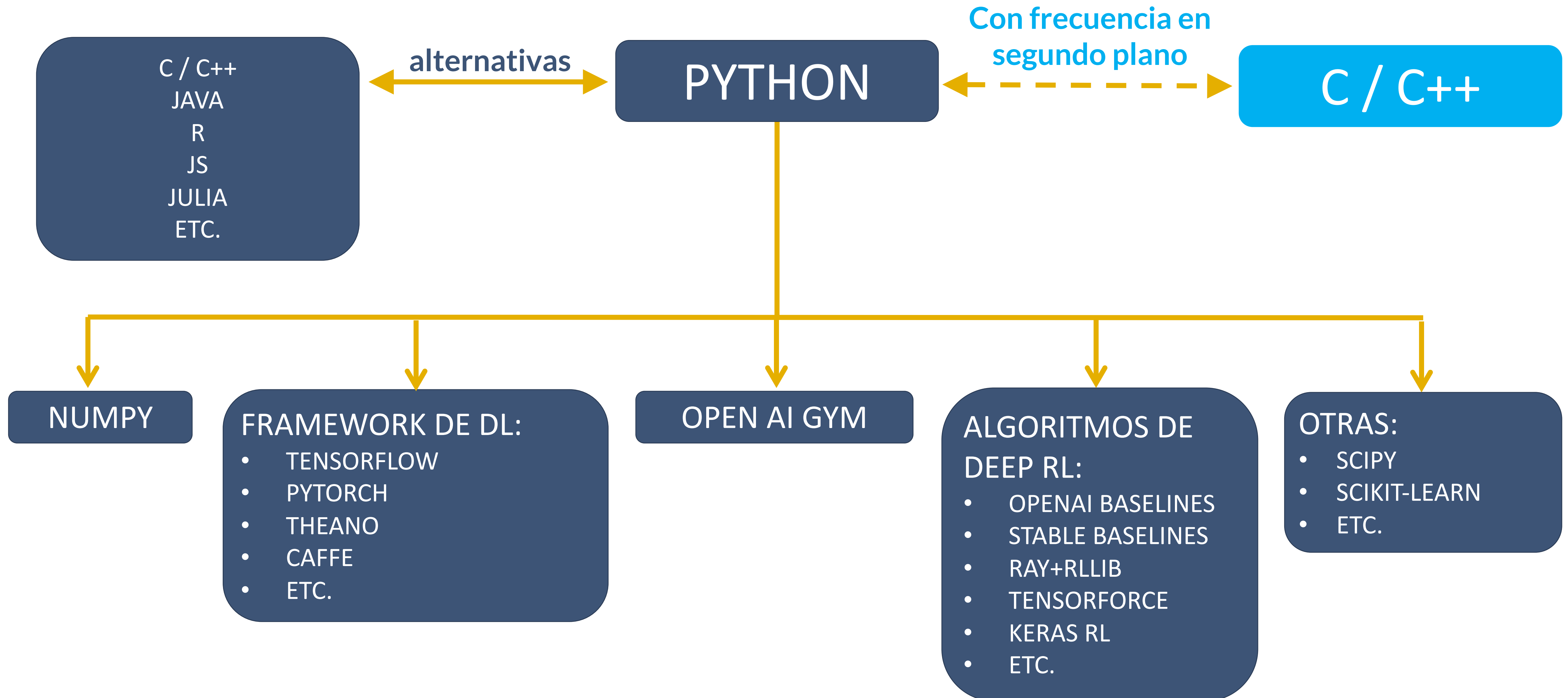
- Neural Information Processing Systems Conference (NIPS)
- International Conference on Learning Representations (ICLR)
- International Conference on Machine Learning (ICML)

TUTORIALES

- Thomas Simonini's Deep Reinforcement Learning Course

Más recursos [aquí](#)

LENGUAJES Y LIBRERÍAS





JUGADORES CLAVE

EMPRESAS

[Bhalla \(2017\). An Introduction to Deep Reinforcement Learning](#)

JUGADORES CLAVE

INVESTIGADORES

- David Silver, Vlad Mnih & **Oriol Vinyals** (Google DeepMind)
- Peter Abeel & Sergey Levine (BAIR, UC Berkeley)
- Richard Sutton (University of Alberta)
- Ilya Sutskever, Rocky Duan & John Schulman (OpenAI)
- Marc Deisenroth & Jan Peters (Imperial College London & Max Planck Institute for Intelligent Systems respectivamente)
- Jason Weston (Facebook AI Research)

Bhalla (2017)

Sergey Levine & Peter Abeel from UC Berkeley
Imagen de [The New York Times](#)





Saturdays.AI
Murcia



UCAM
UNIVERSIDAD
CATÓLICA DE MURCIA



GRACIAS. ¿PREGUNTAS?

INTRODUCCIÓN AL APRENDIZAJE POR REFUERZO

Prof. Antonio Serrano Fernández

<https://antonioserranofz.com/>