# Introducción

**Sistemas Inteligentes Distribuidos**

Sergio Alvarez
Javier Vázquez

# Inteligencia artificial: ¿qué definición?

*The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.*

(Oxford Dictionary of English)

# Inteligencia artificial: ¿qué definición?

*Artificial intelligence is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.*

*The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.*

(Encyclopaedia Britannica)

# Perspectiva multidisciplinar

- **Filosofía**
  - ¿Es posible usar reglas para generar conclusiones válidas?
  - ¿Cómo emerge la mente de un sistema físico como el cerebro?
  - ¿De dónde procede el conocimiento?
  - ¿Puede el conocimiento generar acción o agencia? ¿Cómo?
- **Lógica**
  - ¿Cuáles son las reglas formales que nos permiten generar conclusiones válidas?
  - ¿Qué es computable y qué no?
  - ¿Cómo podemos razonar a partir de conocimiento incompleto o estocástico?

# Perspectiva multidisciplinar

- **Neurociencia**
  - ¿Cómo procesa la información el cerebro?
  - ¿Es posible formalizar e implementar este proceso?
- **Psicología**
  - ¿Cómo piensan y actúan los humanos? ¿Y los animales, individual o colectivamente?
- **Economía/Sociología**
  - ¿Cómo surge o emerge un comportamiento colectivo (social, organizacional) a partir del comportamiento individual?
- **Teoría de control**
  - ¿Cómo se comportan los artefactos que se pueden auto-controlar?

# Tipos de inteligencia artificial

| | |
|---|---|
| **Think like people** | **Think rationally** |
| **Act like people** | **Act rationally** |

# Una definición más global

*Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals*.

*AI-based systems can be purely software-based, acting in the virtual world or AI can be embedded in hardware devices.*

(European Commission's Communication on AI)

# Una definición más global

*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal.*

*AI systems can* ***either use symbolic rules or learn a numeric model,*** *and they can also* ***adapt their behaviour by analysing how the environment is affected by their previous actions***.

(High-Level Expert Group on AI)

https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

# Preguntas abiertas: SID

¿Qué implica que un sistema de IA esté situado en un entorno?
¿Cómo se puede adaptar a, o aprender de, los cambios en el entorno?

¿Cómo se consigue que la IA pueda cumplir con objetivos complejos?
¿Cómo se modelan los objetivos y se traducen en una toma de decisiones?

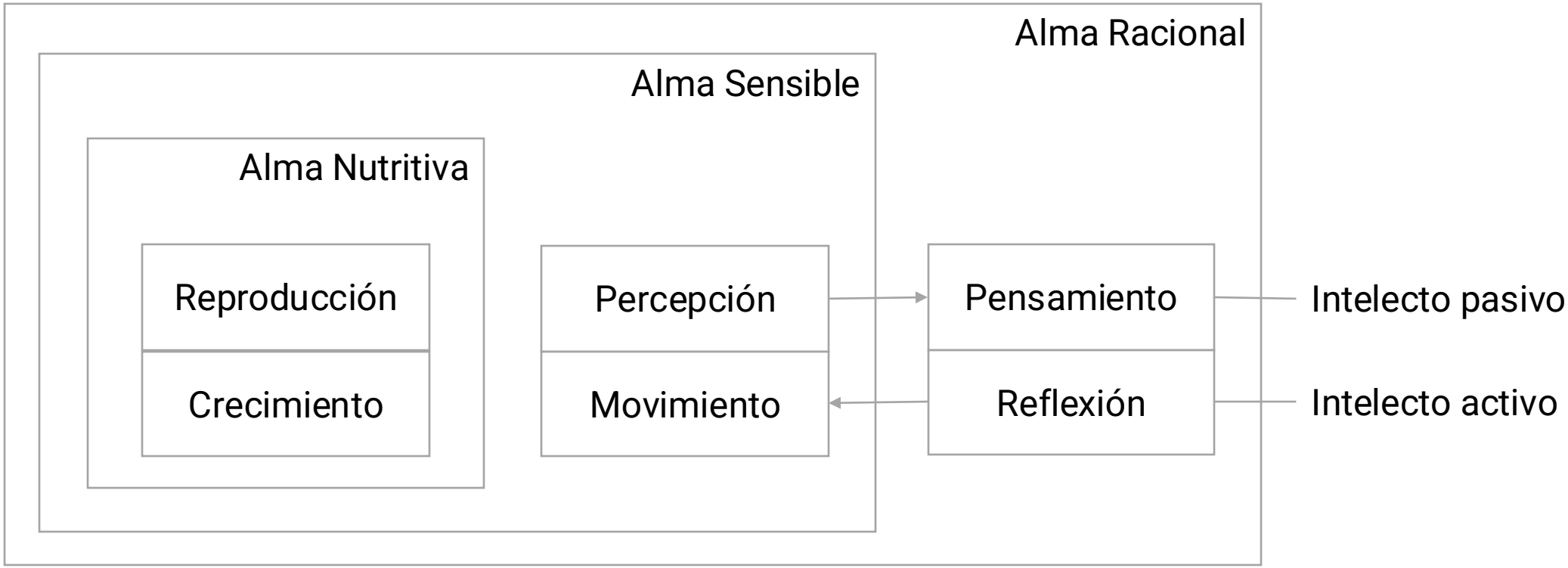
¿Puede un sistema de IA ser simbólico y subsimbólico a la vez?
¿Es posible combinar diferentes paradigmas en un mismo sistema?
¿Qué ocurre si el entorno incluye otros sistemas inteligentes?

# Una aproximación: cognición

- **Obtener inspiración en el concepto de cognición**
  - Facultad de procesar información a partir de la percepción y las experiencias para generar conocimiento
  - Este conocimiento adquirido permite descubrir nuevas inferencias, tomar decisiones, planificar, aprender y comunicarse
- **Diferentes metáforas**
  - Lógica
  - Búsqueda en un espacio de estados
  - Razonamiento basado en el conocimiento: reglas, patrones, experiencias
  - Sistemas evolutivos
  - Sistemas sociales o socio-técnicos
  - Aprendizaje estadístico
  - Conexionismo

# Arquitecturas cognitivas

Aristóteles (*De Anima*, c. 350 a.C.*)*

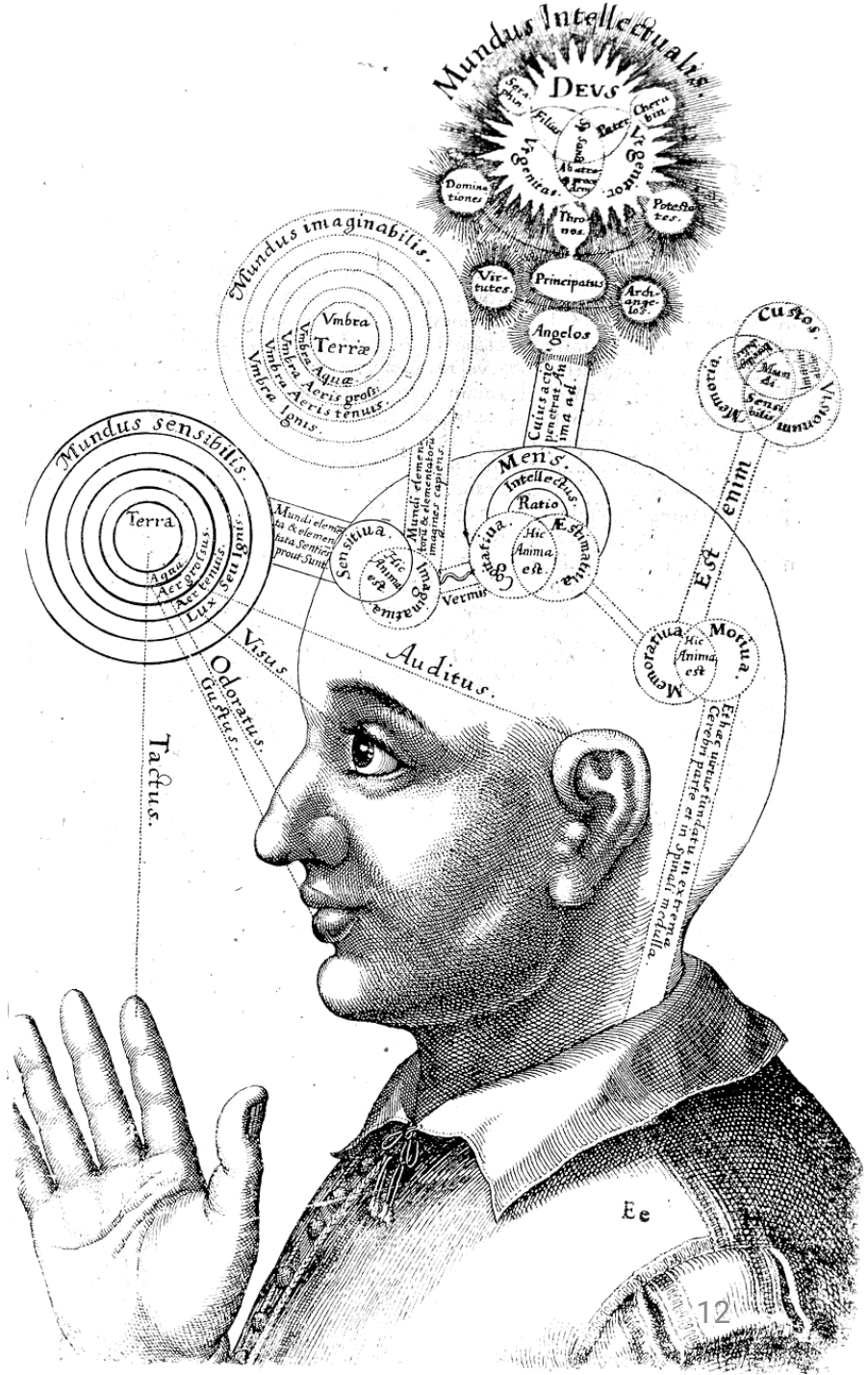


# Arquitecturas cognitivas

Robert Fludd (*Tomus Secundus,* c. s. 1617)

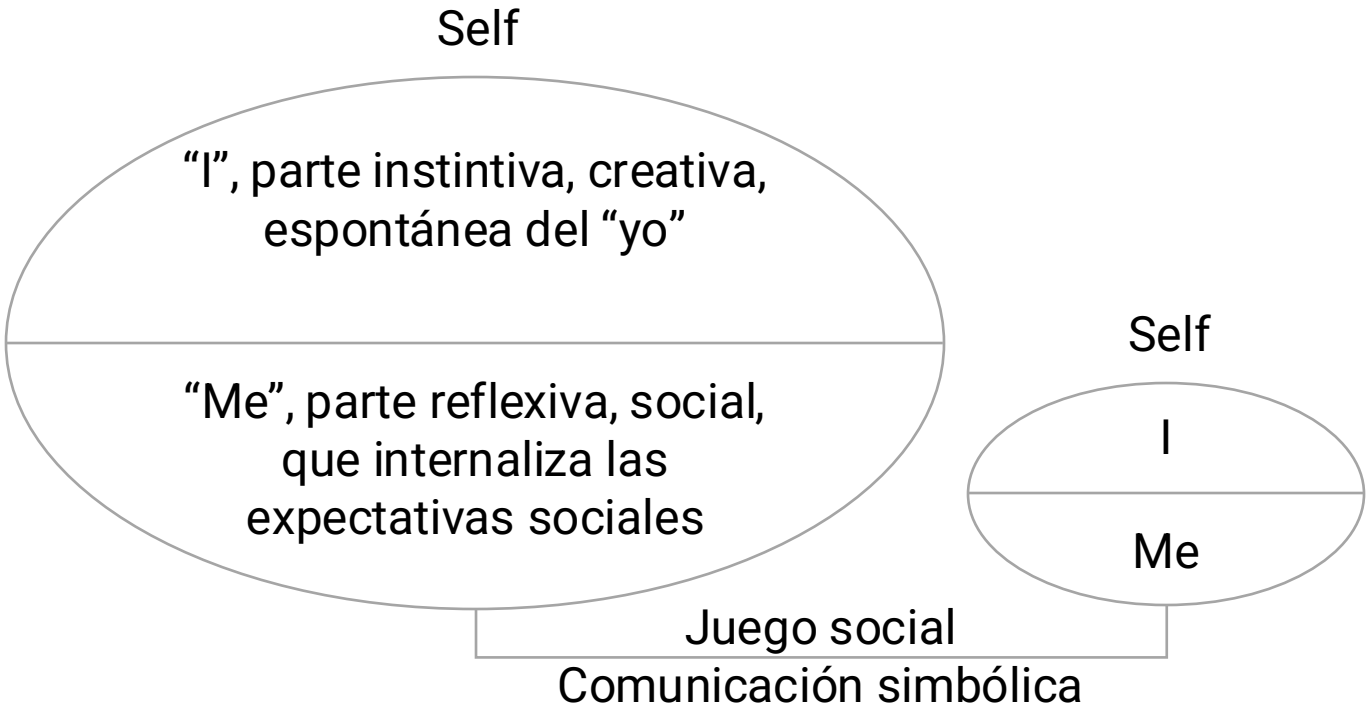
Mezcla de evidencias científicas, teología e ideas esotéricas

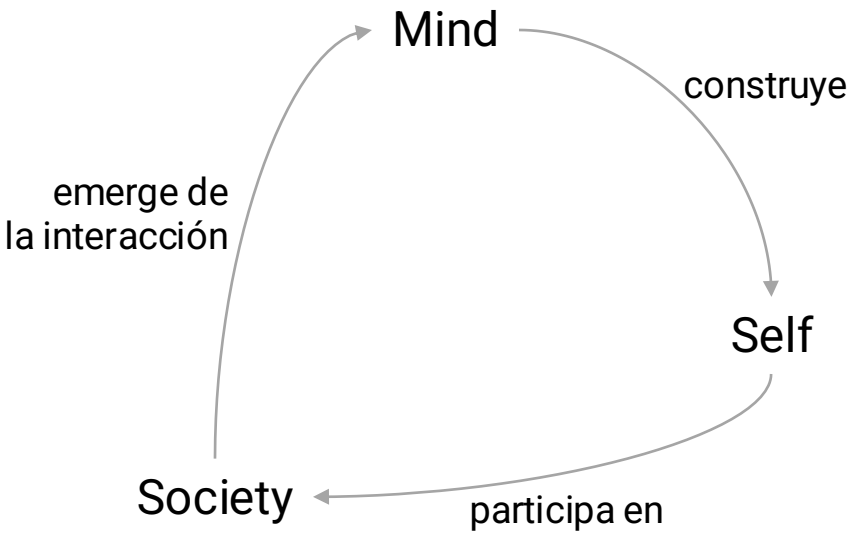
- *Mundus Intellectualis*, conexión mente-microcosmos. El pensamiento se materializa por acción divina.
- *Mundus Imaginibilis*, desconectado de influencias externas
- *Mundus Sensibilis*, conectado a la percepción



# Arquitecturas cognitivas

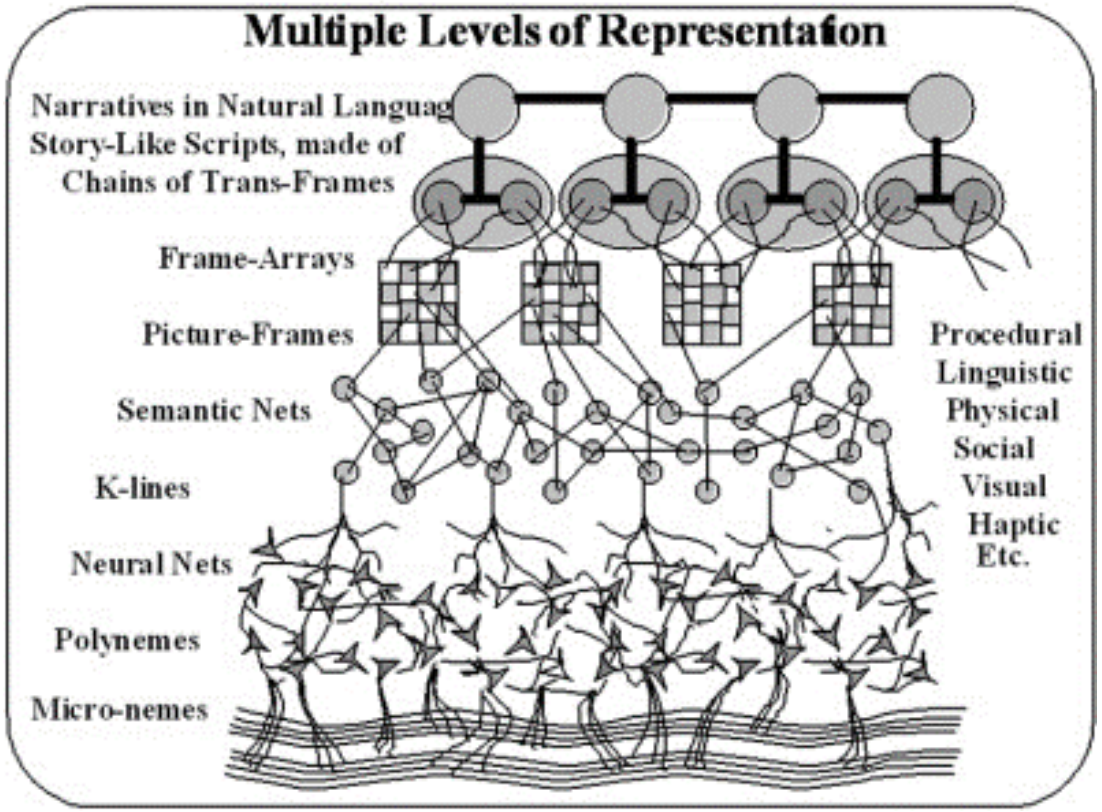
George Herbert Mead (*Mind, Self, and Society*, c. 1863-1931)



# Arquitecturas cognitivas

Marvin Minsky (*The Society of Mind, The Emotion Machine*, c. 1980)

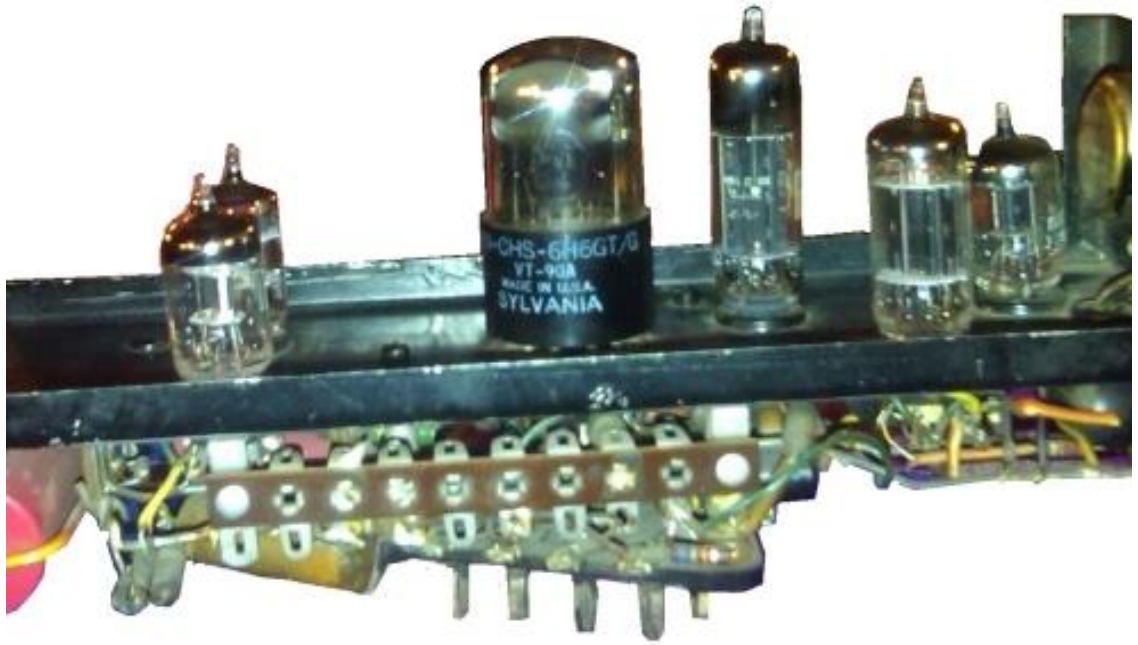



*We shall envision the mind (or brain) as composed of many partially autonomous* ***"agents"—as a "Society" of smaller minds***

*Think of the brain as 400 different computers or databases […] The idea of this theory is that* ***you have a system which switches among different mental states***

# Arquitecturas cognitivas

Marvin Minsky (***SNARC***, c. 1951)




***Stochastic Neural Analog Reinforcement Computer***

Hardware capaz de aprender a partir de 40 memorias (*neuronas*) conectadas por sinapsis

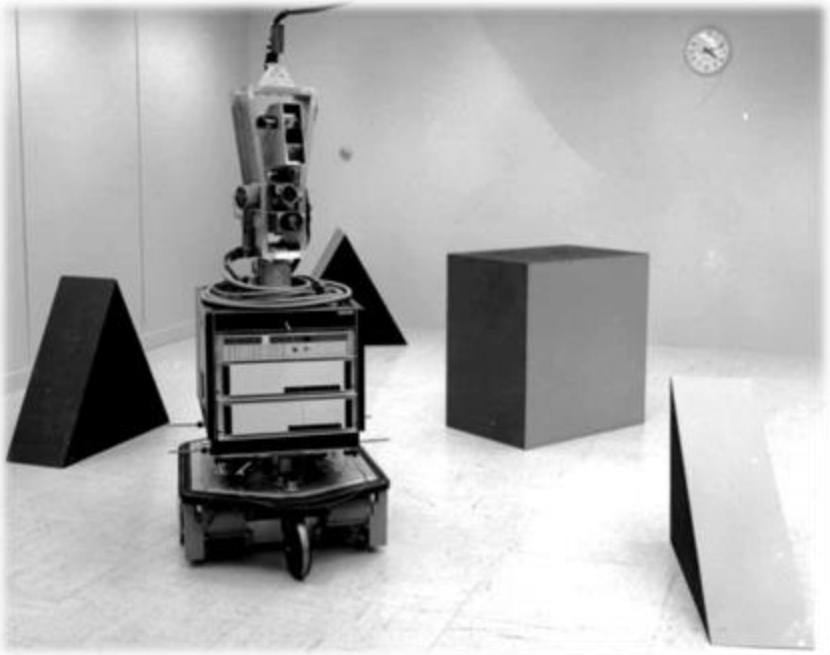
La máquina disponía de memoria sobre el pasado para adaptarse y funcionar de manera eficiente frente a situaciones diversas

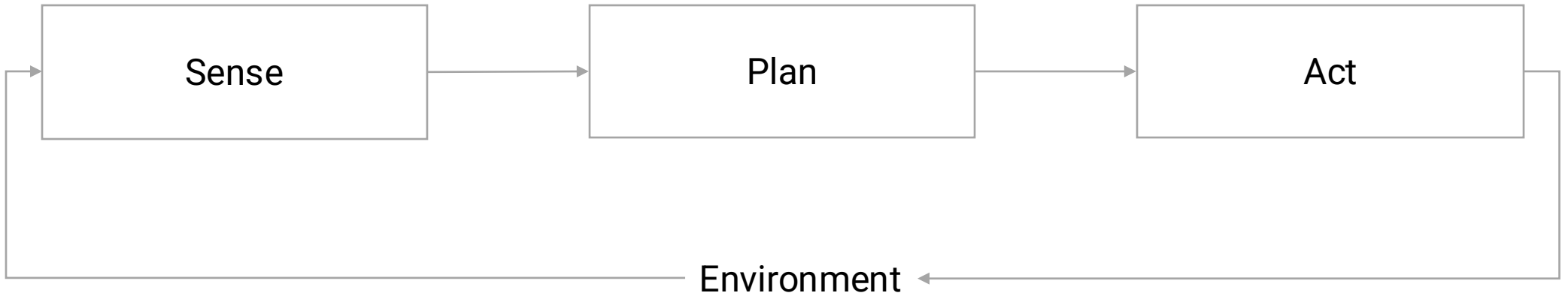
# Arquitecturas cognitivas

**Sense-Plan-Act** (1967-?)

Con la robótica, aparecieron enfoques prácticos y puramente reactivos

Enfoque simbólico, por ejemplo con Shakey implementado con STRIPS (Stanford)



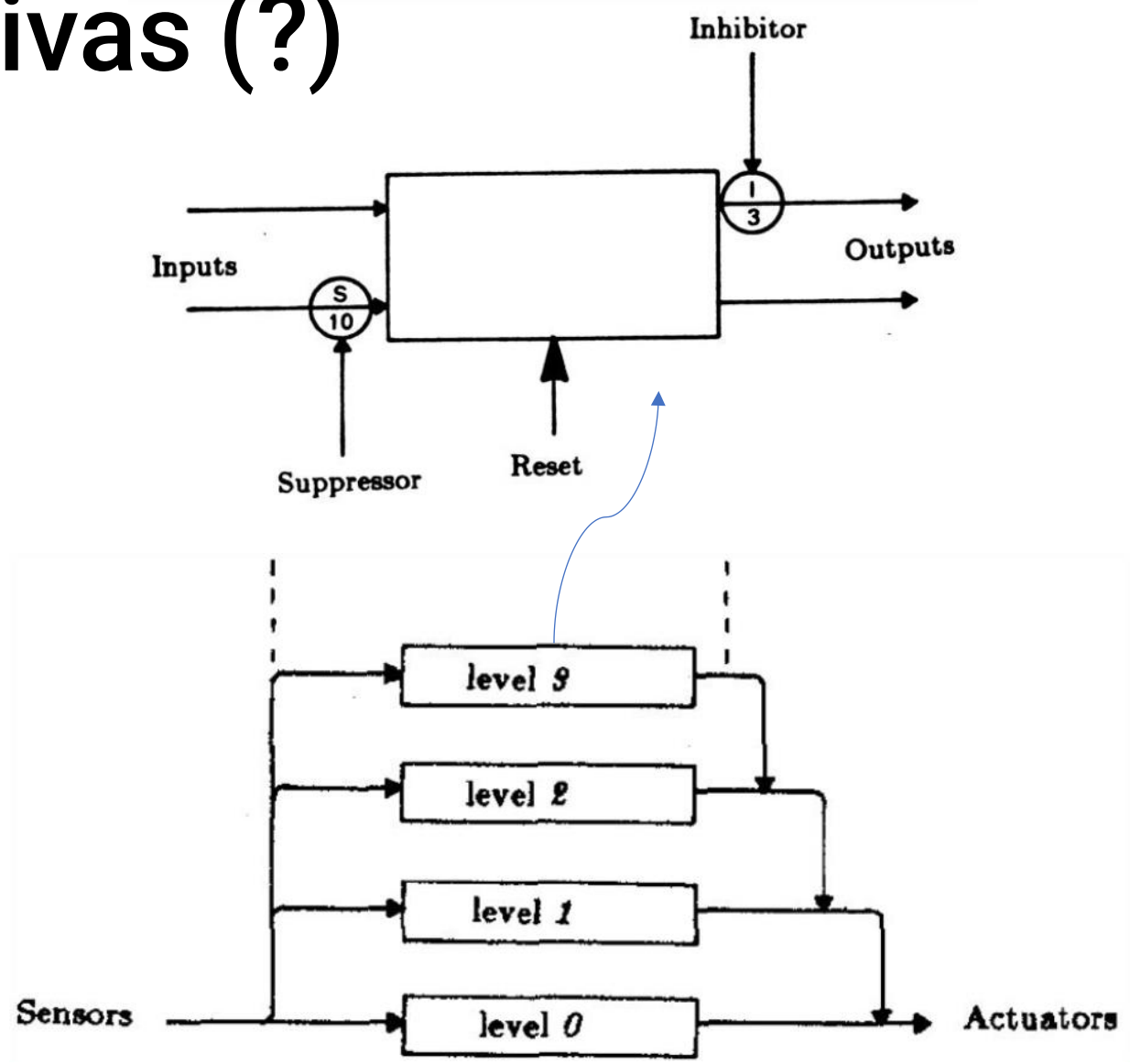


# Arquitecturas cognitivas (?)

Rodney Brooks (**Subsumption**: *Elephants don't play chess*, 1990)

**Hipótesis:** no es necesario ningún tipo de representación simbólica para tener inteligencia artificial

Acoplamiento directo sensores – actuadores a partir de módulos, a diferentes niveles de abstracción, implementados como FSMs

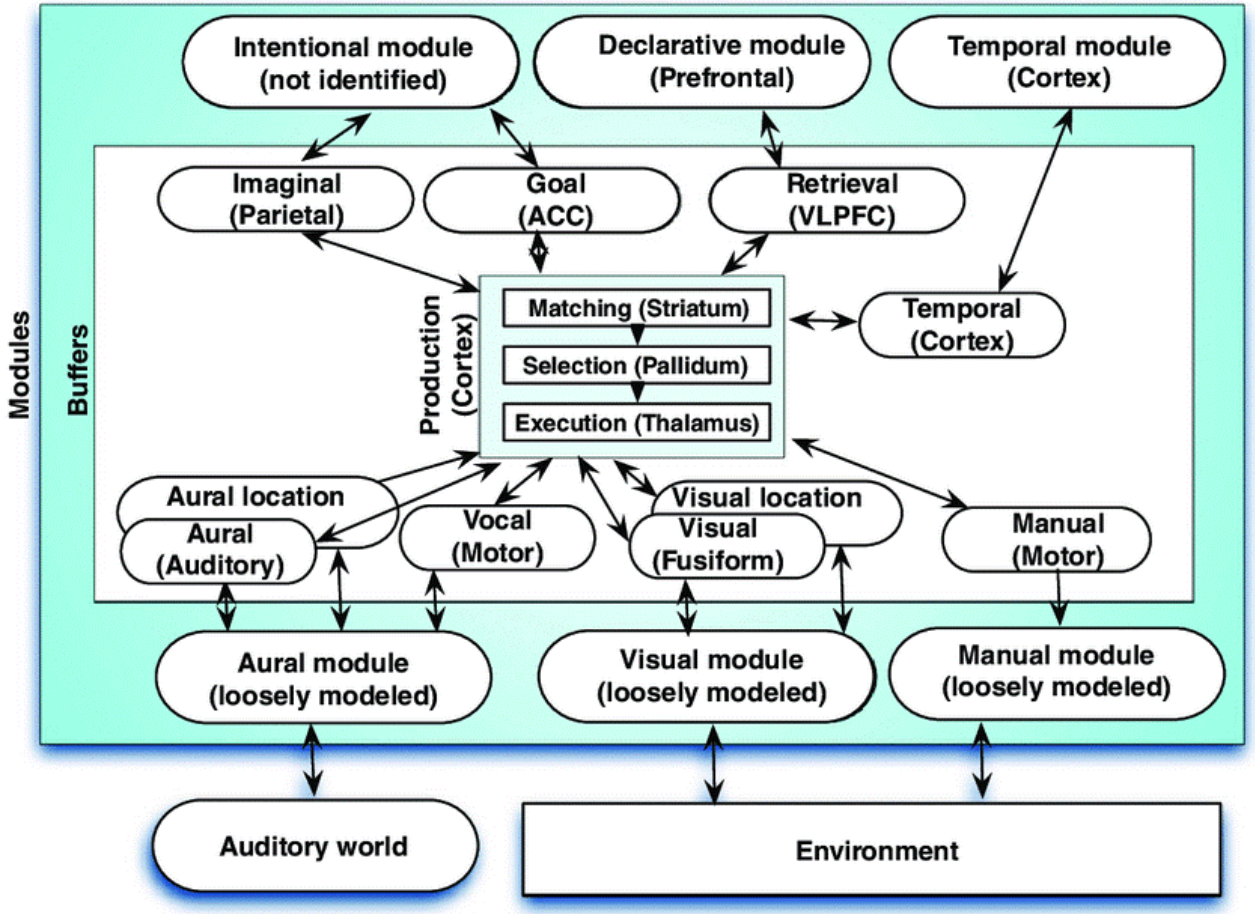


# Arquitecturas cognitivas

Inspiración psicológica: **ACT-R** (1973-?)

Modelo cognitivo basado en diversas teorías sobre el pensamiento y el aprendizaje humanos

Hipótesis: la mente opera a través de módulos especializados (visión, movimiento, memoria) a partir de símbolos y reglas dinámicas



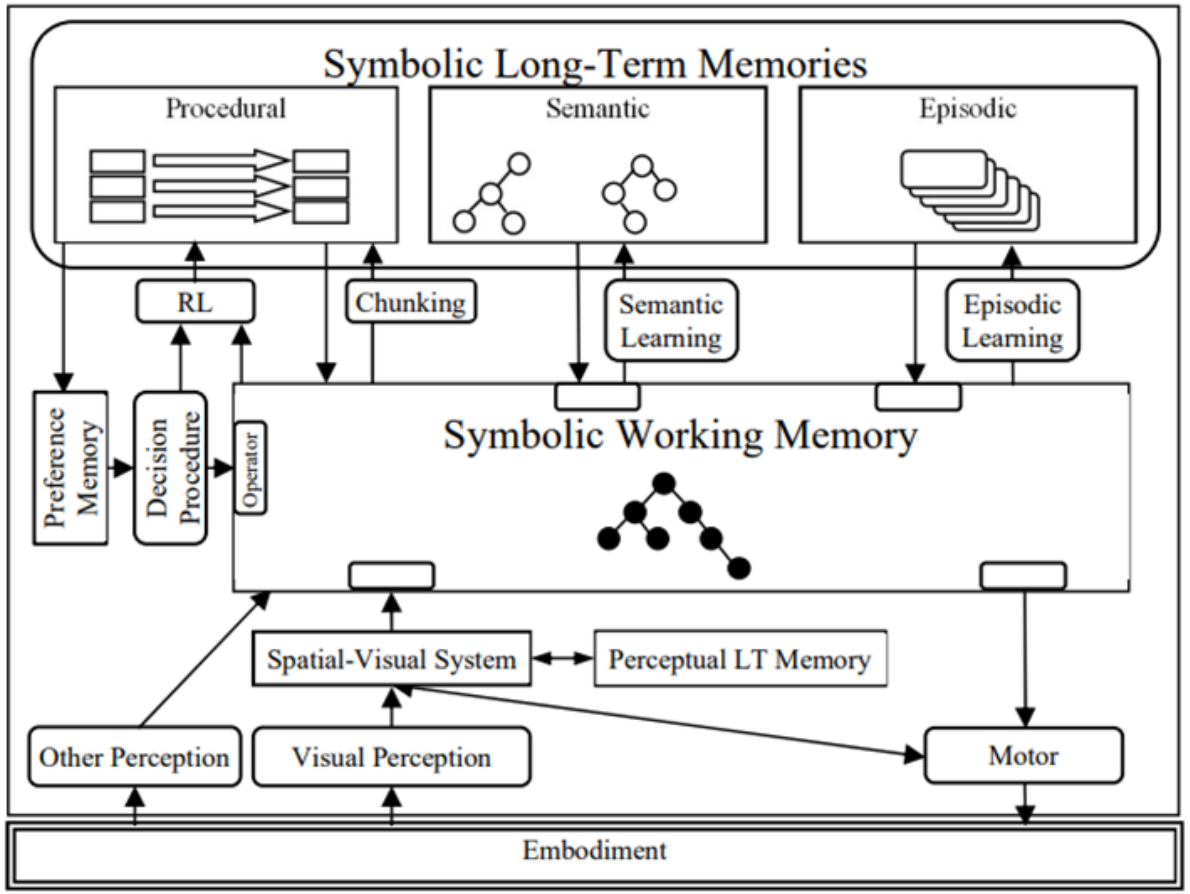

Ritter, Frank E., Farnaz Tehranchi, and Jacob D. Oury. "ACT-R: A cognitive architecture for modeling cognition." *Wiley Interdisciplinary Reviews: Cognitive Science* 10, no. 3 (2019): e1488.

# Arquitecturas cognitivas

Inspiración computacional:
**SOAR/Sigma** (1983-?)

Modelos cognitivos que intentan agrupar todos los tipos de cognición natural y artificial en una **arquitectura genérica y general**

Cada módulo es un sistema probabilístico de producción que evoluciona con el aprendizaje



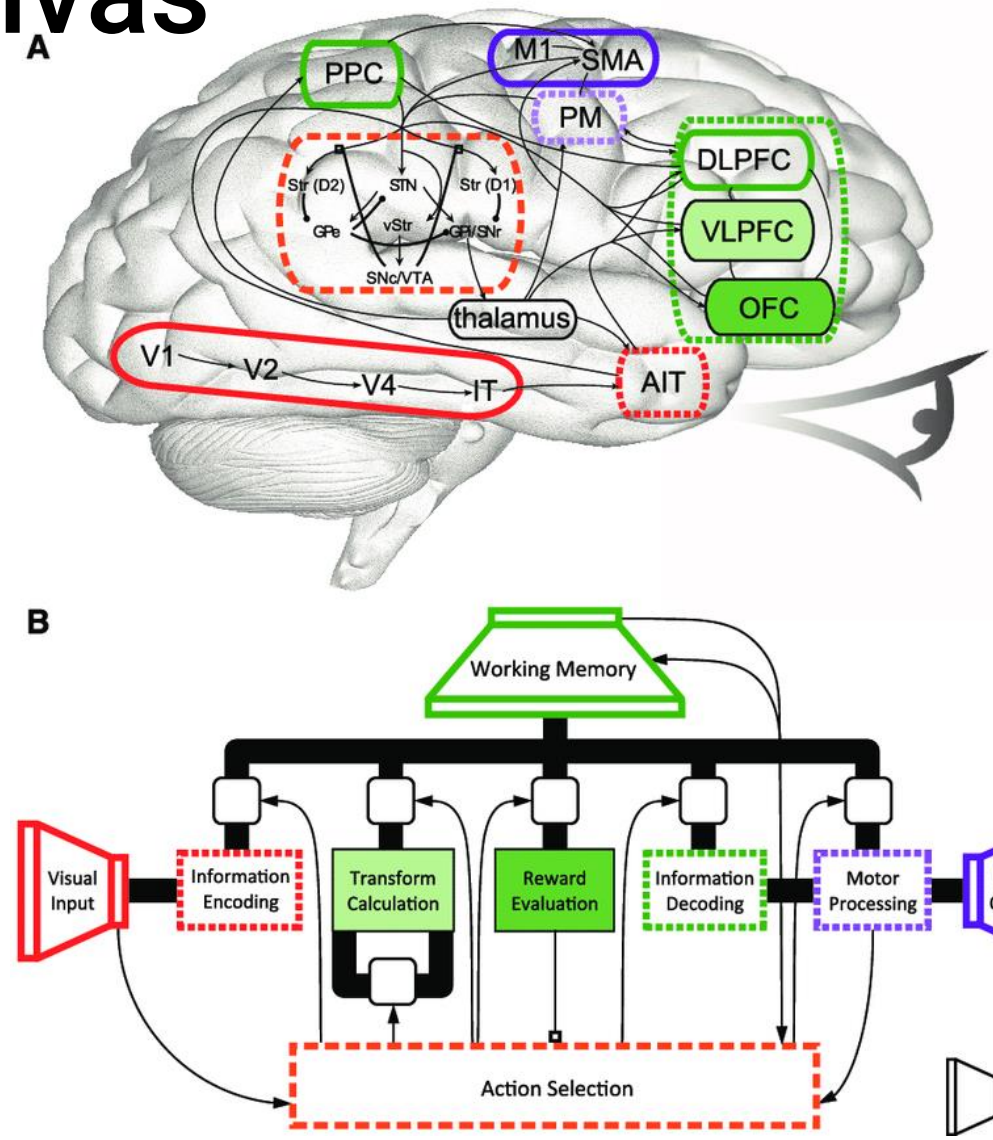
# Arquitecturas cognitivas

Inspiración biológica: **SPAUN** (2012-?)

Modelo del cerebro animal, permite simulaciones a gran escala

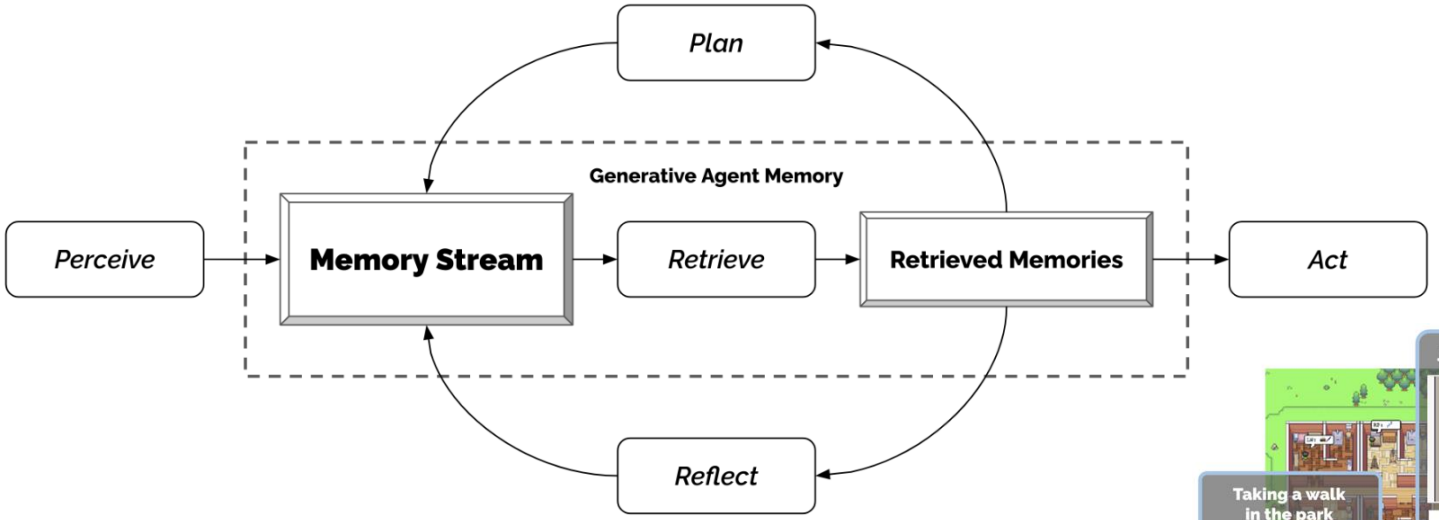
Teoría de los punteros semánticos: el cerebro codifica explícitamente representaciones intencionales de alto nivel (simbólicas)

2.5 millones de neuronas, visión artificial y un brazo robótico



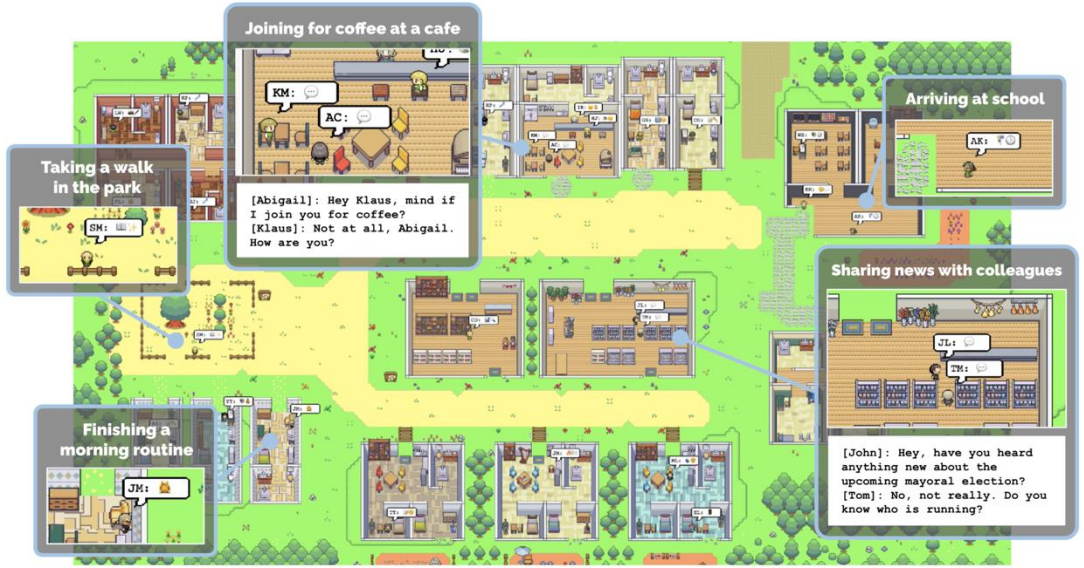

Stewart, T., Choo, F. X., & Eliasmith, C. (2012). Spaun: A perception-cognition-action model using spiking neurons. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).

# Arquitecturas cognitivas




Hay un creciente interés en la actualidad por construir sistemas inteligentes capaces de actuar, usando modelos grandes de lenguaje (*LLM agents*)

Sin embargo, se está detectando la necesidad de complementar el modelo de lenguaje con componentes que permitan tener memoria a corto/largo plazo, priorizar conocimiento/tareas, planificar, o tener un *world model*



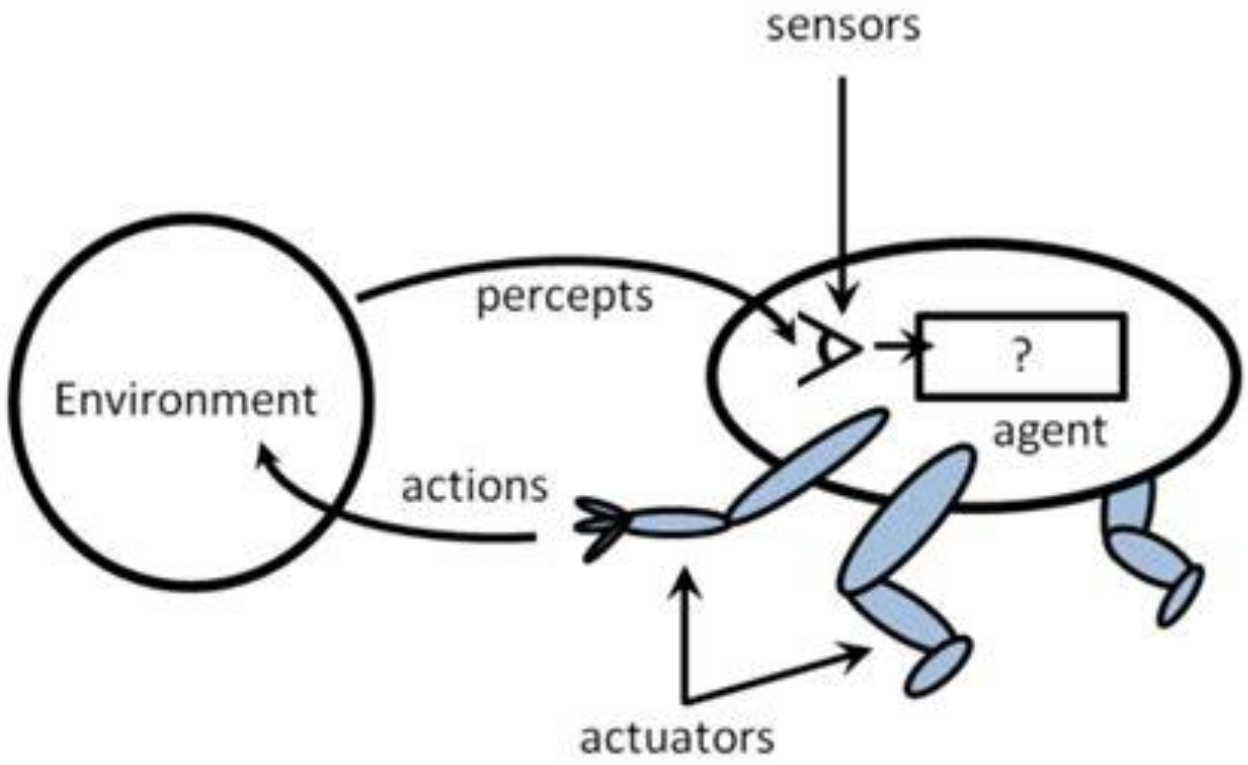

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29–November 01, 2023, San Francisco, CA, USA*. ACM, New York, NY, USA 22 Pages. https://doi.org/10.1145/3586183.3606763

# Arquitecturas cognitivas

- ¿Hay arquitecturas mejores que otras?
  - Parámetros: generalidad, versatilidad, racionalidad, capacidad de aprendizaje, escalabilidad, reactividad, eficiencia, alineamiento
- ¿Llegaremos a definir una arquitectura cognitiva completa y general?
- ¿Qué tipo de inteligencia podemos obtener con una arquitectura cognitiva?

# Paradigma de agente inteligente

- Russell & Norvig (*AIMA*)
- Abstracción suficientemente genérica
- Un agente es un sistema computacional al que se le presupone capacidad de acción autónoma en algún entorno para cumplir con unos objetivos previamente diseñados
- Un agente es capaz de percibir su entorno y actual en él



# Paradigma de agente inteligente

- Utilizaremos esta abstracción para explorar cómo se construyen artefactos (de software) que exhiban comportamiento *racional* **(Mitad 1)**
  - **Pregunta principal:** ¿cómo construir un sistema que, con cierto grado de autonomía, produzca un resultado correcto dada una entrada?
- Más adelante veremos cómo componer múltiples instancias de estas abstracciones para obtener sistemas inteligentes distribuidos **(Mitad 2)**
  - **Pregunta principal:** ¿cómo construir un sistema que opere con cierto grado de autonomía y de manera estable en un entorno distribuido, potencialmente de gran escala, y donde posiblemente existan otros sistemas que no poseemos o no controlamos?

# Dos escenarios

**Internet of Things**

Computación y comunicación ubicua, interfaces inteligentes

Entornos con números arbitrarios de sistemas distribuidos móviles y embebidos, interactuando en nombre de usuarios

Características clave: autonomía, distribución, adaptación, tiempo de respuesta

**Edge computing**

Computación y datos a diferentes niveles para acercarse al usuario:

- Edge: *wearables*, móviles, sensores
- Fog: subestaciones eléctricas, antenas, routers urbanos
- Cloud: instalaciones dedicadas de gran escala

Características clave: eficiencia, seguridad, privacidad, latencia

# Objetivos del curso (guía docente)

- Proveer a los estudiantes los conocimientos matemáticos y computacionales suficientes para analizar sistemas distribuidos inteligentes mediante modelos adecuados
  - Sistemas inteligentes: sistemas lógico-simbólicos (BDI, ontologías, lógica modal), sistemas subsimbólicos (procesos de decisión de Markov, aprendizaje por refuerzo)
  - Sistemas distribuidos: teoría de juegos, sistemas socio-técnicos
- Ilustrar diversas estrategias de coordinación y mostrar cómo implementarlas y optimizarlas
  - Cooperación: elección social, diseño de mecanismos, consenso, normas e instituciones
  - Competición: diseño de mecanismos, juegos adversariales

# Temas interesantes que no veremos

- Formación de coaliciones
- Búsqueda cooperativa de caminos
- Planificación distribuida
- Swarm intelligence
- Web semántica
- Reputación y confianza
- Emociones
- …

# Estructura del curso

# Estructura del curso

AGENTES INTELIGENTES

Tipos de entorno
Arquitecturas de agente: reactivo, deliberativo

Agentes dirigidos por objetivos
Agentes BDI
Lógica modal
Lógica descriptiva y ontologías

Agentes dirigidos por utilidad
Procesos de decisión de Markov (MDPs)
Aprendizaje por refuerzo

**Objetivos prácticos:**

Conocer el estado del arte en arquitecturas de agente
Crear agentes deliberativos usando formalismos lógicos
Entender las diferencias entre los distintos tipos de lógica modal
Implementar y razonar con ontologías, usando axiomas de la lógica descriptiva
Obtener políticas óptimas para MDPs
Entrenar agentes usando algoritmos de aprendizaje por refuerzo

# Estructura del curso

SISTEMAS MULTI-AGENTE

Teoría de juegos: cooperación, competición
Equilibrio de Nash
MDPs descentralizados

Diseño de mecanismos: subastas
Elección social
Algoritmos de consenso

Búsqueda Monte Carlo
Aprendizaje por refuerzo multi-agente

Sistemas sociotécnicos
Normas, organizaciones, valores
Ética, alineamiento y transparencia

**Objetivos prácticos:**

Calcular equilibrios de Nash algorítmicamente

Resolver escenarios de elección social

Entrenar sistemas multi-agente con algoritmos de aprendizaje por refuerzo

Entender cómo se puede alinear un sistema multi-agente a unos valores propuestos

Aplicar técnicas de coordinación a agentes BDI *(extra)*

# Evaluación

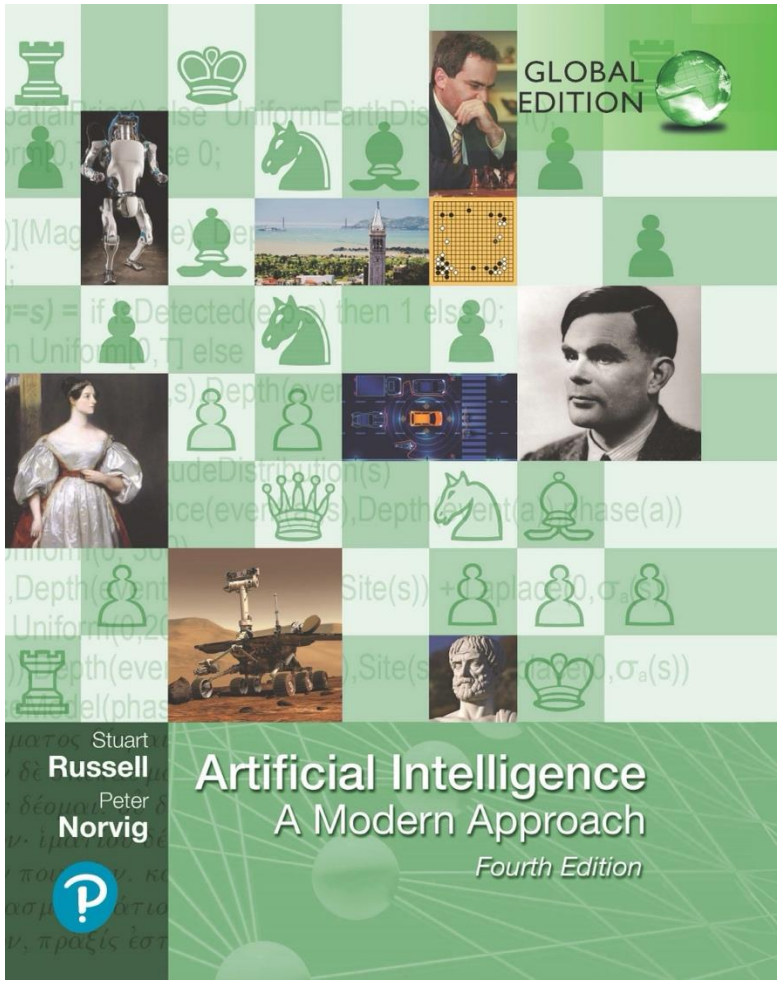
- Nota de teoría (T): 50%
  - 50% examen parcial, 50% examen final
  - La nota del parcial es recuperable en el final (sin tiempo adicional)
- Nota de problemas (P): 20%
  - Cuatro problemas sencillos (mismo peso) que se propondrán en el Racó cada dos semanas aproximadamente
  - Algunos serán de investigación de literatura existente, otros serán problemas prácticos que resolver
  - Algunas sesiones de laboratorio serán de problemas
- Nota de laboratorio (L): 30%
  - Tres prácticas (mismo peso: agentes BDI, aprendizaje por refuerzo, aprendizaje por refuerzo multi-agente)
  - Habrá opción a puntos extra
- Para P y L, tenéis que formar grupos de tres personas
  - Se aceptarán grupos de 2 o de 4 de manera **excepcional** y **fundamentada**
- Los contenidos se renovaron hace un año, **no lo baséis todo en los exámenes y prácticas que encontréis publicados**

# Calendario

| | Tema T | Teoría | Sesión L | Laboratorio | Entregas |
|---|---|---|---|---|---|
| **11/2** | 1 | Introducción | 1 | Agentes (I) | |
| **18/2** | 2 | Agente inteligente | 2 | Agentes (II) | |
| **25/2** | 3 | Agentes lógico-simbólicos | 3 | Agentes (III) | P1 (Arquitecturas) |
| **4/3** | 4 | Ontologías | 4 | Ontologías | |
| **11/3** | 5 | Agentes guiados por utilidad | 5 | Problemas: MDPs | P2 (Ontologías) |
| **18/3** | 6 | Aprendizaje por refuerzo | 6 | Aprendizaje por refuerzo (I) | |
| **25/3** | 6 | Aprendizaje por refuerzo | 7 | Aprendizaje por refuerzo (II) | L1 (Agentes) + P3 (MDPs) |
| **1/4** | 7 | Teoría de juegos | - | PARCIALES | |
| **8/4** | - | PARCIALES | 8 | Problemas: teoría de juegos | |
| **15/4** | - | FESTIVO | - | FESTIVO | |
| **22/4** | 8 | Coaliciones, Elección social y consenso | 9 | Problemas: cooperación | L2 (Aprendizaje por refuerzo) |
| **29/4** | 9 | Competición, búsqueda Monte Carlo | 10 | Problemas: competición | |
| **6/5** | 10 | Diseño de mecanismos | 11 | Búsqueda Monte Carlo | |
| **13/5** | 11 | Aprendizaje por refuerzo multi-agente | 12 | Aprendizaje por refuerzo multi-agente (I) | P4 (Teoría de juegos) |
| **20/5** | 12 | Sistemas socio-técnicos | 13 | Aprendizaje por refuerzo multi-agente (II) | |
| **27/5** | 13 | Estado del arte, IA responsable | 14 | Aprendizaje por refuerzo multi-agente (III) | |
| **3/6** | - | EXAMEN | - | - | L3 (Aprendizaje por refuerzo multi-agente) |

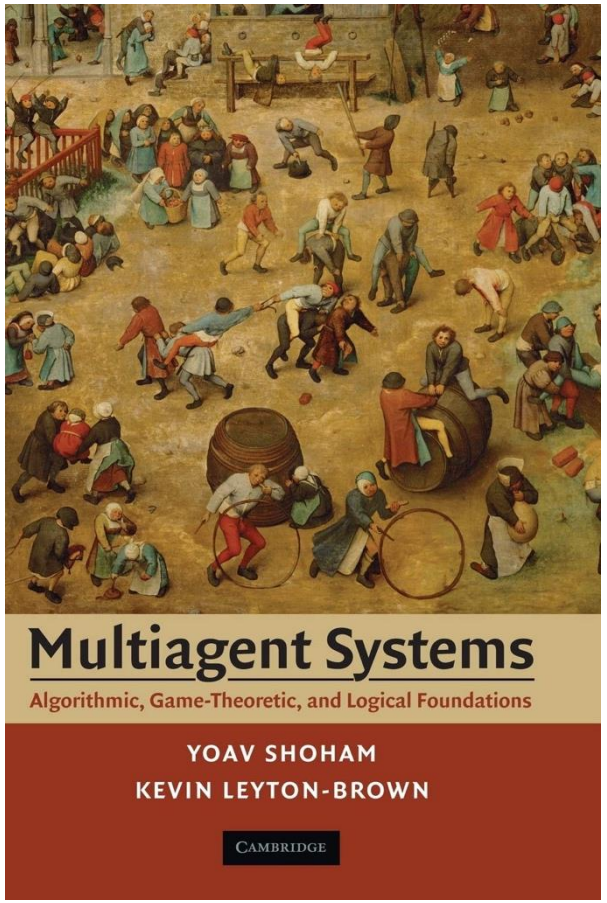
# Bibliografía seguida en el curso

- *Artificial Intelligence: A Modern Approach 4th edition, Global edition* (Russell & Norvig)
- Ediciones anteriores no sirven
- Capítulos 2, 6, 7, 10, 16, 17, 23, 28
- Hay un par de copias en el campus, pero si no lo encontráis, avisadnos



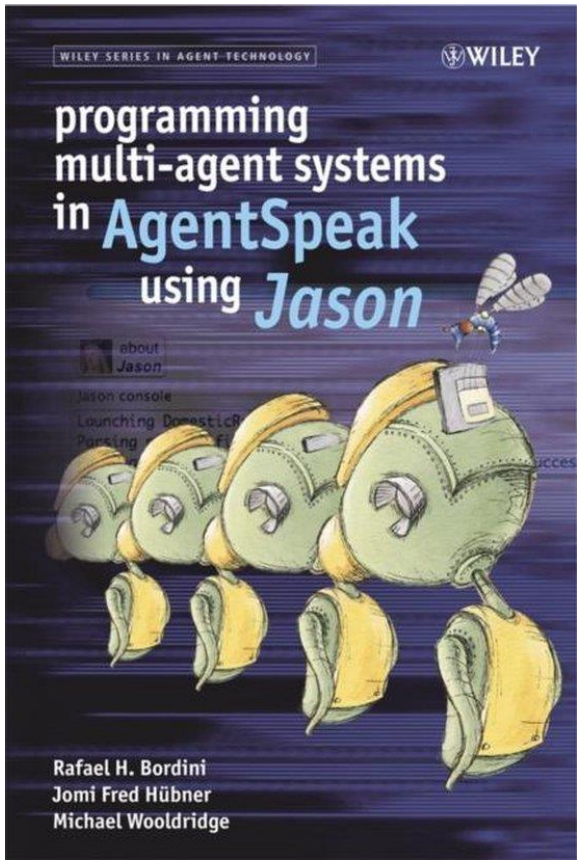
# Bibliografía seguida en el curso

- *Multiagent systems : algorithmic, game-theoretic, and logical foundations* (Shoham)
- Disponible en http://www.masfoundations.org/mas.pdf
- Capítulos 3, 4, 7, 8, 9, 10, 11, 13, 14



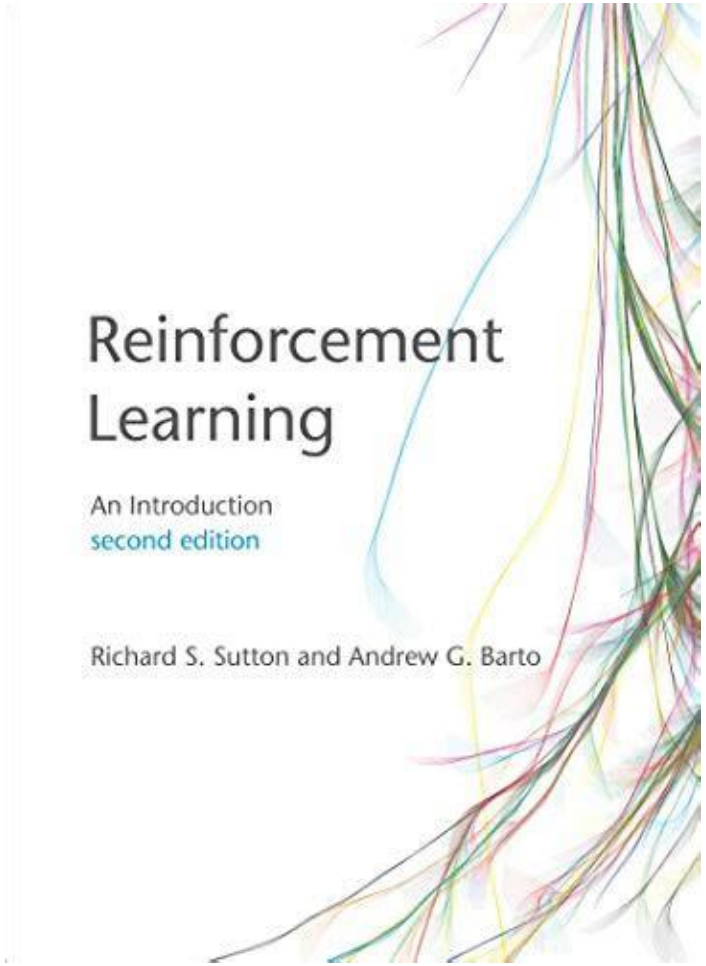
# Bibliografía seguida en el curso

- *Programming multi-agent systems in AgentSpeak using Jason* (Bordini, Hübner & Wooldridge)
- Hay un par de copias en la biblioteca, si no lo encontráis, avisadnos
- Capítulos 2, 3, 10



# Bibliografía seguida en el curso

- *Reinforcement learning: an introduction* (Sutton & Barto)
- Hay una copia en la biblioteca, si no lo encontráis, avisadnos
- Capítulos 3, 5, 6



# Bibliografía seguida en el curso

- *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches* (Albrecht, Chistianos & Schäfer)
- Disponible en https://www.marl-book.com/
- Capítulos 2, 3, 4, 5, 6