



# Tema 1

## “Introducción a la Inteligencia Artificial”

*Año Académico 2019/20*

*Profesores:*

*Alberto Fernández y Holger Billhardt*

# Tema 1: Introducción a la IA

## Resumen:

### 1. Introducción a la Inteligencia Artificial

#### 1.1 Qué es la IA?

#### 1.2 Agentes Inteligentes

# Inteligencia Artificial

**Objetivo:** *estudiar* los entes inteligentes

- científico: *entender* los entes inteligentes
- ingenieril: *construir* entes inteligentes

**Algunas definiciones:**

- Sistemas que *piensan* como *humanos*

“La interesante tarea de lograr que las computadoras piensen...  
Máquinas con mente, en su amplio sentido literal” (Haugeland 1985)

- Sistemas que *actúan* como *humanos*

“El arte de crear máquinas con capacidad de realizar funciones que realizadas por personas requieren inteligencia” (Kurzweil 1990)

- Sistemas que *actúan* de forma *racional*

“La rama de la Informática que se ocupa de la automatización del comportamiento inteligente” (Luger & Stubblefield, 1993)

“IA fuerte”



“IA débil”

# Pensar como humanos

## Modelado cognitivo:

- abrir la “caja negra” de la mente humana
- analizar los procesos mentales (introspección, experimentos)
- desarrollar una teoría acerca de los procesos mentales
- aplicar esta teoría en la simulación de dichos procesos en un ordenador

## Ejemplo. General Problem Solver (GPS) [Newell & Simon 1961]:

- resuelve problemas mediante la descomposición en subproblemas más simples
- se centra en la comparación de los pasos de razonamiento del GPS con los pasos seguidos por una persona al resolver el mismo problema

## Ciencia Cognitiva:

- modelos computacionales (IA) + técnicas experimentales (psicología)
- construir teorías rigurosas y verificables acerca de los procesos mentales

# Actuar como humanos

## Prueba de Turing : [Alan Turing, 1950]

- Un evaluador humano y un interlocutor están separados por una mampara
- El interlocutor puede ser bien otra persona o bien un ordenador
- El evaluador formula preguntas a través de un teletipo, y el interlocutor da sus respuestas del mismo modo
- El ordenador supera la prueba, si el evaluador no es capaz de distinguir entre él y un humano

## Capacidades requeridas :

- procesamiento del lenguaje natural
- representación del conocimiento y razonamiento
- Aprendizaje automático

## Prueba total de Turing:

- incluye señales de vídeo y objetos físicos
- requiere capacidad de visión computacional y robótica

# Actuar de forma racional

## Racionalidad:

- prescriptivo: cómo las personas *deberían* actuar
- sentido estricto: ¿cómo sacar “conclusiones verdaderas”?
- sentido amplio: ¿cómo actuar y “sobrevivir” en un entorno?

## Pensar de forma racional:

- leyes de pensamiento de Aristóteles: razonamiento irrefutable
- lógica formal :
  - lenguaje formal para representar todo tipo de entes en el mundo
  - modelo riguroso para razonar sobre dichos entes
- en su estado “puro”, más estrechamente relacionado con la filosofía y las matemáticas

## Actuar de forma racional:

- Inteligencia Artificial: modelar/construir sistemas que *actúan* basándose en la inferencia lógica automática

# Actuar de forma racional

## Agentes Racionales:

- enfoque relativo al contexto: actuar de forma correcta en un *entorno*
- no se limita a la inferencia racional (lógica)
  - a veces es imposible determinar formalmente cuál es la mejor acción
  - en algunas situaciones es “racional” emprender una acción “buena” inmediatamente en vez de esperar hasta determinar la alternativa óptima
- se pueden determinar acciones eficientes por inferencias no lógicas

# Áreas en IA

- Búsqueda y Planificación
- Representación del conocimiento y Razonamiento
- Aprendizaje automático
- Sistemas multi-agente
- Robótica
- Percepción computacional
  - visión, procesamiento de lenguaje natural
- Internet
  - análisis de redes sociales, crowdsourcing

# Áreas investigación “de moda”

- Aprendizaje automático a gran escala
- Deep Learning
- Aprendizaje por refuerzo
- Robótica: entrenar robot, interactuar con personas
- Visión computacional
- Procesamiento de lenguaje natural: reconocimiento, dialogar, traducción entre idiomas
- Sistemas colaborativos: autónomos colaboran con otros y con humanos
- Crowdsourcing y computación humana
- Teoría de juegos y elección social
- Internet de las cosas (IoT)

# Aplicaciones IA presente/futuro

- Transporte
  - Coches inteligentes, vehículos autónomos, distribución autónoma, planificación, transporte bajo demanda
- Robots caseros/servicios
  - Limpieza, reparto paquetes, seguridad
- Asistencia médica
  - Análisis datos, robots (quirúrgicos, mejora de servicios, cuidado mayores)
- Educación
  - Tutores interactivos
- Seguridad pública
  - Cámaras y drones, detección fraude financiero, predicción crimen
- Entretenimiento
  - Redes sociales, creación música, vídeo, juegos

# Tema 1: Introducción a la IA

## Resumen:

### 1. Introducción a la Inteligencia Artificial

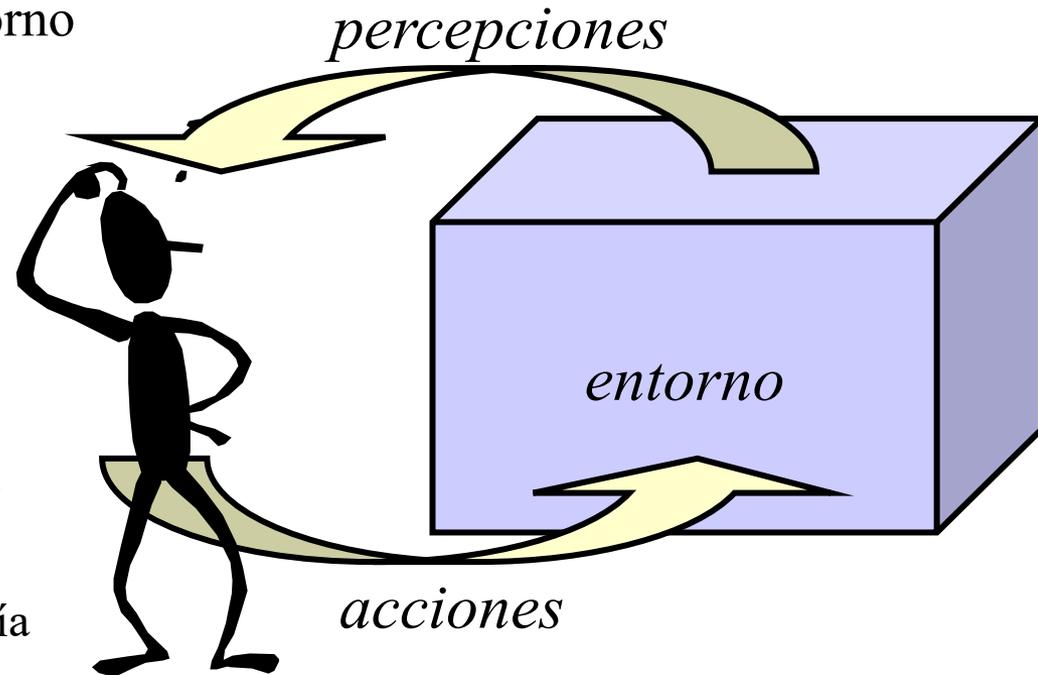
#### 1.1 Qué es la IA?

#### 1.2 Agentes Inteligentes

# Agentes

## Agente:

- ente activo embebido en un entorno
- “cuerpo”:
  - percibe el entorno por medio de *sensores*
  - actúa sobre el entorno por medio de *actuadores*
- “mente”:
  - determina las *acciones* a partir de las *percepciones*
  - *medida de rendimiento* que guía dicho proceso



# Tipos de Agentes

## Agentes naturales:

- cuerpo biológico y entorno natural
- sensores: ojos, oídos, lengua, etc.
- actuadores: piernas, brazos, manos, etc.
- medida de rendimiento: sobrevivir, reproducirse, ...

## Agentes artificiales:

- agentes *hardware* (robots):
  - interactúan directamente con un entorno físico
  - disponen de un “cuerpo” físico
  - sensores: cámaras, telémetros infrarrojos, etc.
  - actuadores: ruedas/piernas, manipuladores, etc.
- agentes *software* (softbots):
  - actúan en entornos virtuales (p.e. Internet)
  - todo software: no necesitan manipular físicamente el entorno
  - sensores y actuadores: dependientes del entorno

# Agente inteligente

## Agentes inteligentes:

- actúan de forma *racional* en su entorno
- *determinantes* de un comportamiento racional :
  - *medida de rendimiento*: define el grado de éxito del agente
  - *secuencia de percepciones*: la experiencia del agente
  - *conocimientos a priori* sobre su entorno
  - *capacidades*: las *acciones* que el agente pueda emprender

## Comportamiento racional:

- a partir de la secuencia de percepciones hasta el momento, y el conocimiento a priori sobre el entorno
- elegir entre las capacidades la acción que maximice la medida de rendimiento

## Racionalidad $\neq$ Omnisciencia

- la selección racional de acciones sólo se basa en la información disponible

# Autonomía

## Problema:

- los conocimientos a priori compilan la “inteligencia” del diseñador
- un agente que no presta atención a sus percepciones
  - no sería inteligente
  - sólo podría actuar en entornos extremadamente simples
  - no puede actuar con éxito en situaciones no anticipadas

## Autonomía:

- “no bajo el control inmediato de una persona”
- un agente es más **autónomo**...
  - ... cuanto más se rige su comportamiento por su propia experiencia
  - ... cuanto menos depende de sus conocimientos a priori

**Agente inteligente = comportamiento racional + autonomía**

# Programa y Arquitectura de Agente

## Programas de Agente:

- software que determina el comportamiento del agente
- implementa la función percepción-acción

Programa agente {

memoria ← *percibir*(memoria, percepción)

acción ← *selección-acción*(memoria, medida-rendimiento)

memoria ← *actuar*(memoria, acción)

}

## Arquitectura de agente:

- los *módulos* que componen el agente
- *estructura* el programa de agente
- partes imprescindibles:
  - componente de percepción
  - componente de selección de acciones
  - componente de acción

# Propiedades del entorno

## Propiedades:

- *accesible* frente a *inaccesible*:
  - ¿El agente puede determinar inequívocamente el estado de su entorno ?
  - Accesible: Ajedrez , “tres en raya”
  - Inaccesible: Póker, laberinto, videojuego
- *determinista* frente a *no determinista*:
  - ¿Las acciones del agente en un estado actual determinan completamente el estado resultante?
  - Determinista: Ajedrez, agente software en entorno simulado
  - No determinista: gestión de tráfico, robot en sala real

# Propiedades del entorno

## Propiedades :

- *estático* frente a *dinámico*:
  - ¿El estado del entorno pueda cambiar mientras que el agente delibera?  
¿Puede cambiar sin que el agente actúe?
  - estático: agente software en un laberinto simulado (entorno no cambia)
  - “semidinámico”: ajedrez (cambios previsibles)
  - dinámico: gestión de tráfico (cambios imprevisibles)
- *discreto* frente a *continuo*:
  - ¿Los conjuntos de posibles percepciones y/o acciones son discretos?
  - discreto: ajedrez, agente software en un laberinto simulado
  - continuo: robot navegando en una sala real