



Universidad
Rey Juan Carlos



Planificación y Sistemas Cognitivos

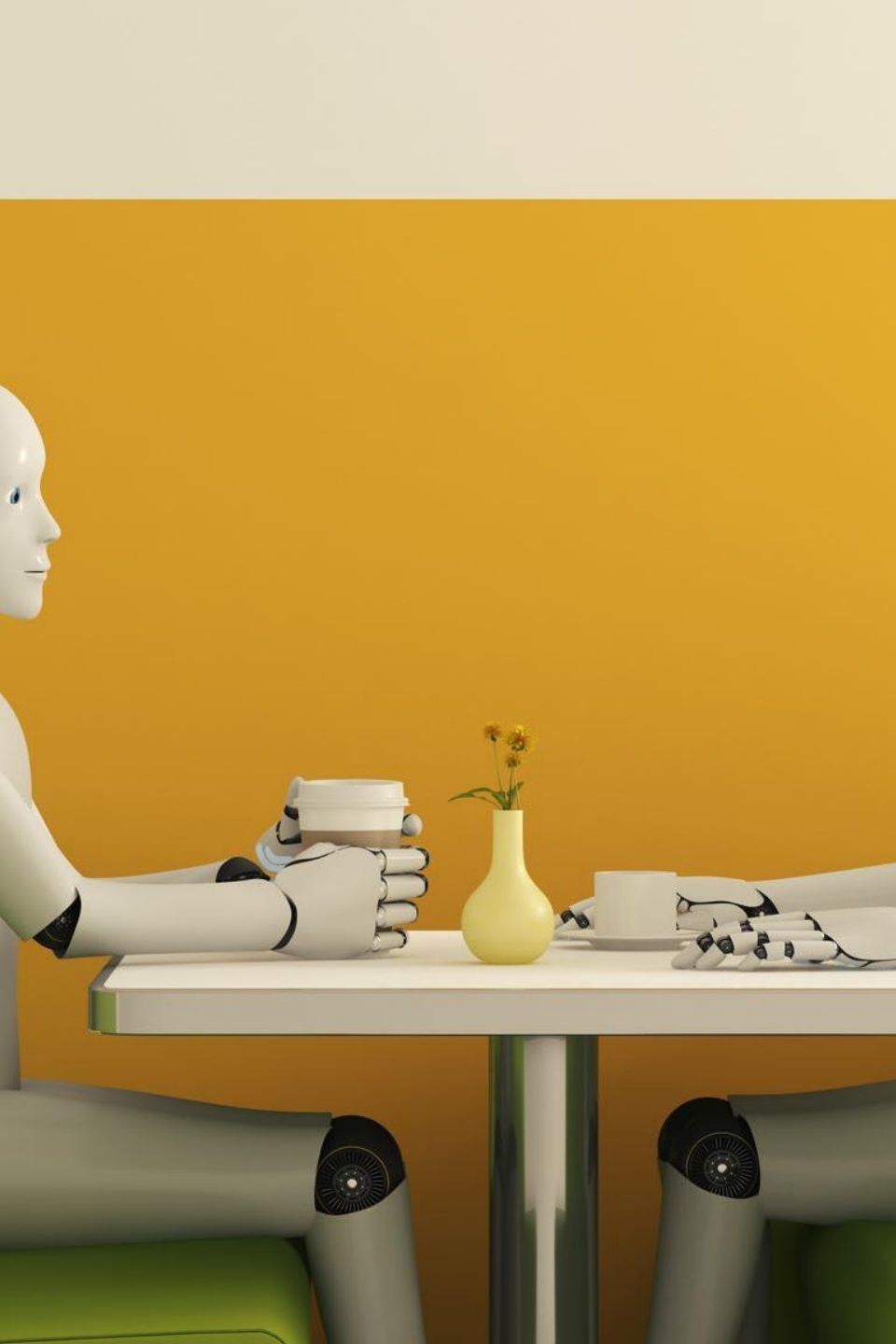
Arquitecturas cognitivas. Aplicaciones

FRANCISCO JOSÉ ROMERO RAMÍREZ

FRANCISCO.ROMERO@URJC.ES

Índice de contenidos

1. Introducción
2. Arquitecturas Cognitivas: Aplicaciones
3. Interacción humano-máquina



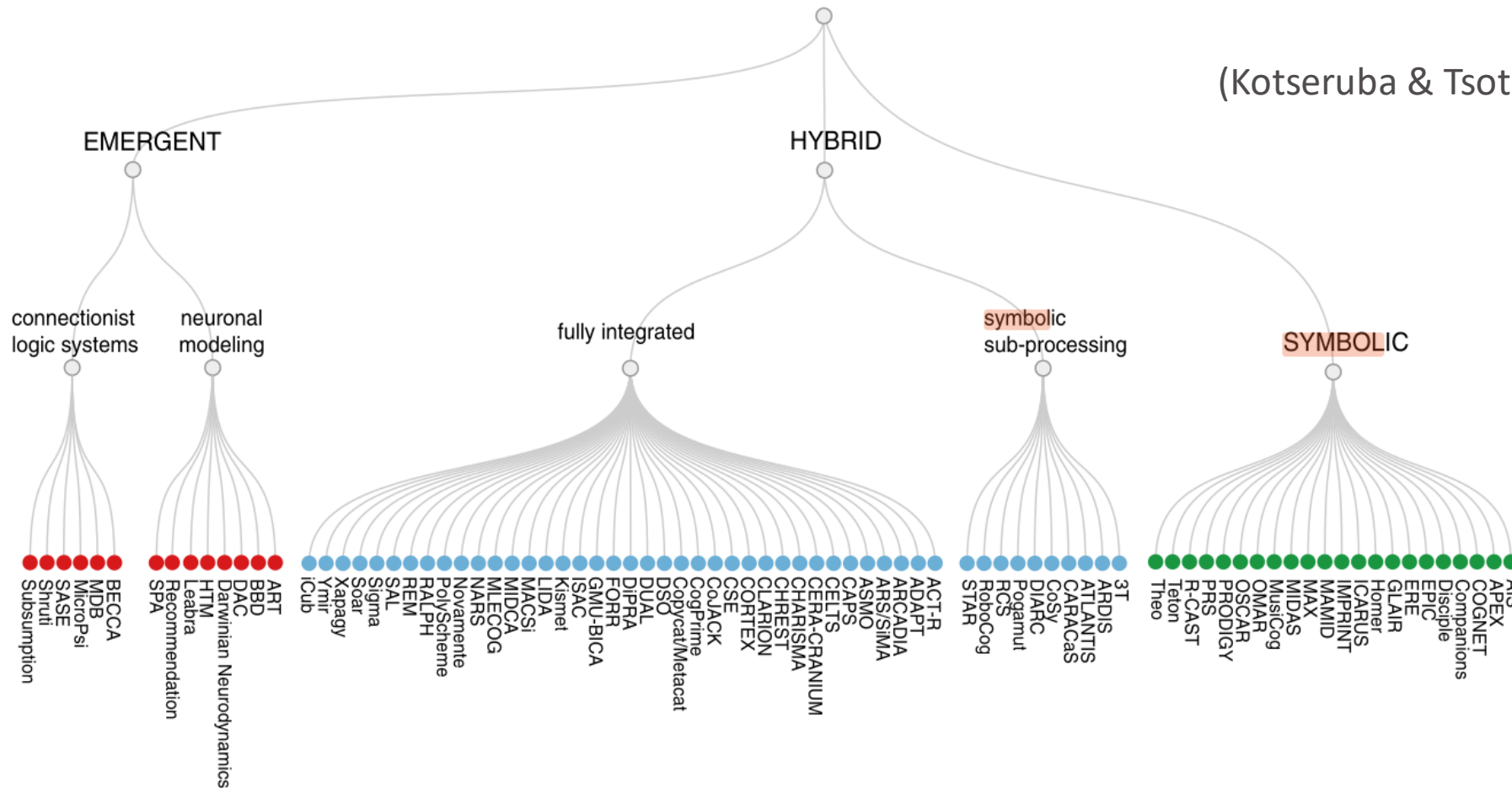
1. Introducción

1. Introducción

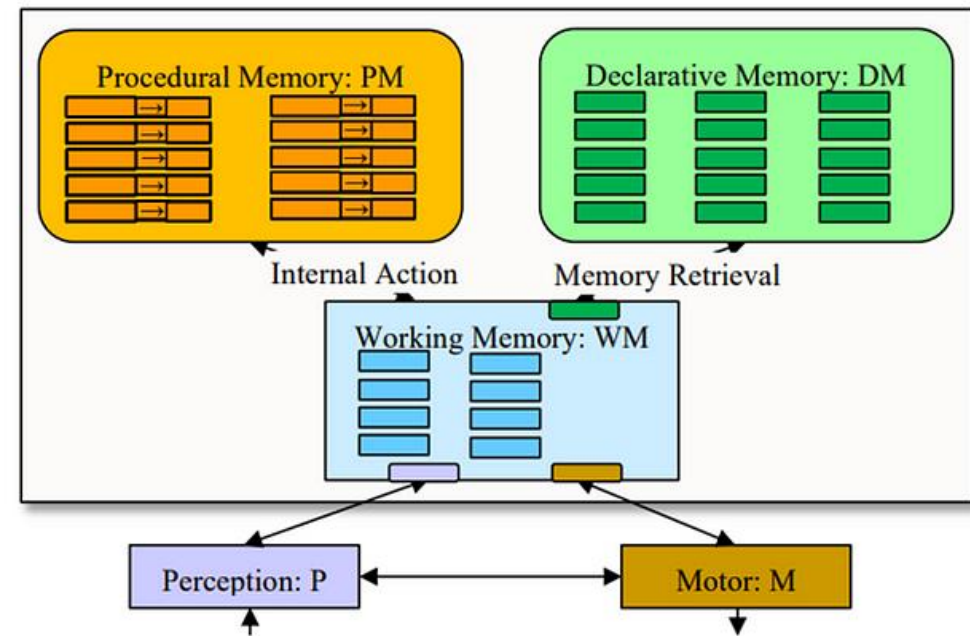
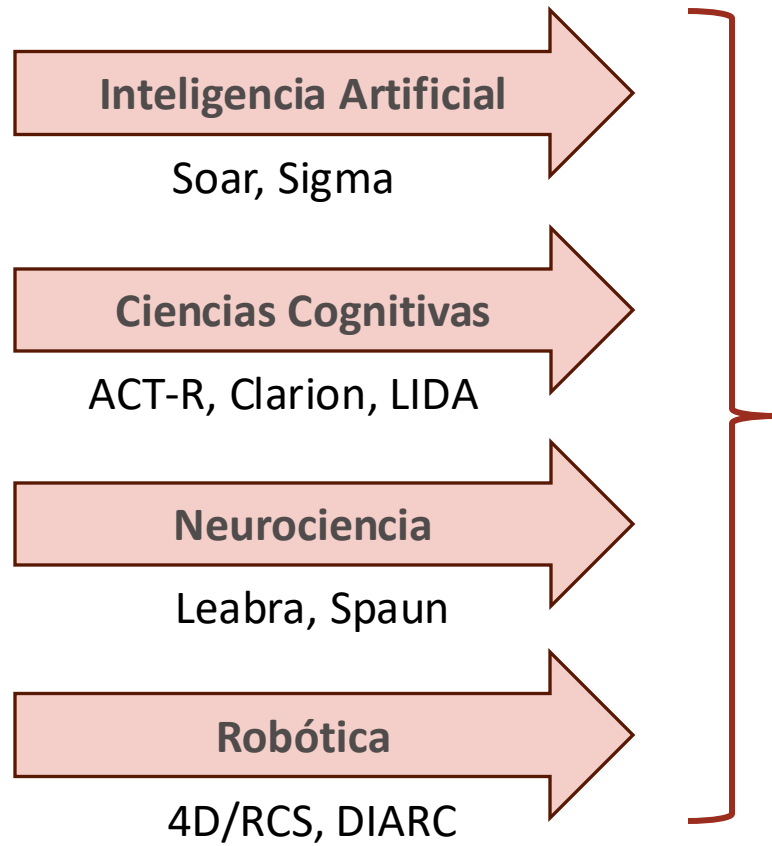
Arquitecturas cognitivas

- Modelo de software inspirado en el funcionamiento de la mente humana.
- Teoría Unificada de la Cognición (A. Newell)
- Que integra capacidades cognitivas: percepción, atención, selección de acciones, memoria, aprendizaje, razonamiento, metacognición, prospección.
- Es común agruparlas de acuerdo a su representación y procesamiento de la información que implementan.
 - Cognitivista (simbólico)
 - Emergente (conexionista)
 - Híbrido
- Utilizando componentes estructurales: **percepción, memoria, razonamiento y acción.**

1. Introducción



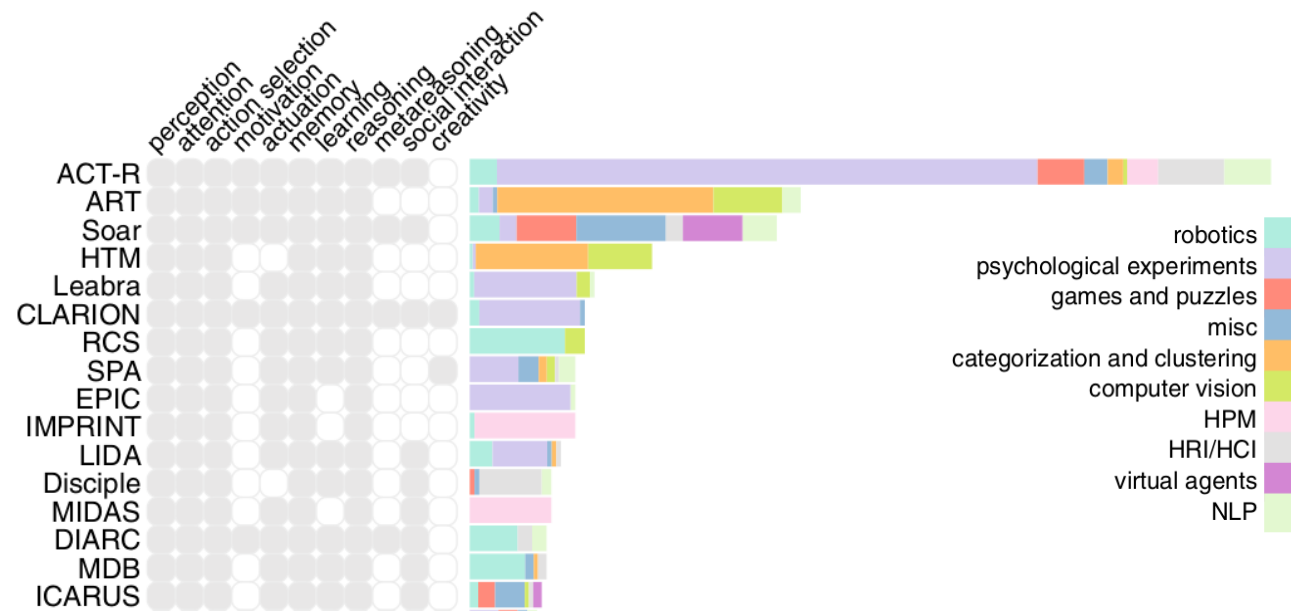
1. Introducción



Modelo común de Arquitectura cognitiva (Laird et al., 2017)

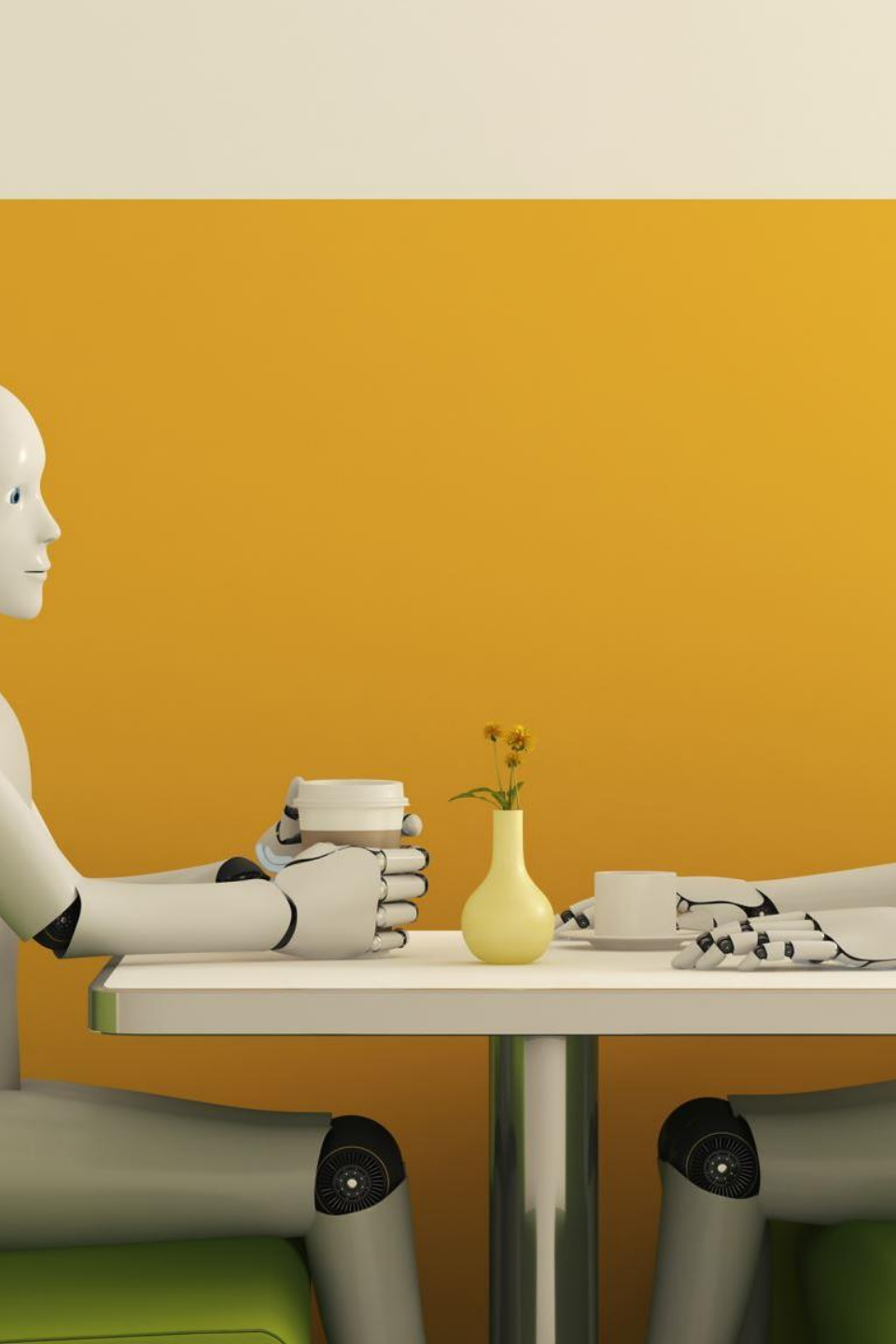
1. Introducción

- (Kotseruba & Tsotsos, 2020) identifican más de 900 proyectos realizados con 84 arquitecturas cognitivas.



http://jtl.lassonde.yorku.ca/project/cognitive_architectures_survey/applications.html

Resumen principales arquitecturas y aplicaciones



2. Arquitecturas cognitivas: Aplicaciones

2. Arquitecturas cognitivas: Aplicaciones

- Principales categorías de aplicaciones prácticas (Kotseruba & Tsotsos, 2020)
 - Modelización del rendimiento humano (HPM)
 - Juegos y rompecabezas
 - **Robótica**
 - Experimentos psicológicos
 - **Procesamiento del lenguaje natural (NLP)**
 - **Interacción persona-robot y persona-ordenador (HRI/HCI)**
 - Visión por ordenador
 - Categorización y agrupación (clustering)
 - Agentes virtuales
 - Otros

2. Arquitecturas cognitivas: Aplicaciones

Robótica

Casi una cuarta parte de las aplicaciones de las arquitecturas cognitivas están relacionadas con la robótica:

1. Navegación y la evitación de obstáculos

- Entornos no estructurados en un vehículo autónomo [RCS (Coombs et al. 2000), ATLANTIS (Gat 1992)]
- Robot móvil [Subsumption (Brooks 1989), CoSy (Pacchierotti et al. 2005)]
- Vehículo marino no tripulado [CARACaS (Huntsberger y Woodward 2011)]

2. Búsqueda y manipulación de objetos

- Agarre de objetos blandos, como peluches [ISAC (Kawamura et al. 2004)]
- Diferentes tipos de agarre (objetos con asas situadas en la parte superior o en un lateral) demostrados en un robot controlado por DIARC (Wilson et al. 2016).
- Adaptar su agarre a latas de diferentes tamaños, cajas y una regla iCub (Sauser et al. 2012).

2. Arquitecturas cognitivas: Aplicaciones

Robótica

3. Pocas arquitecturas implementan múltiples habilidades para escenarios complejos:
 - Vendedor robótico [CORTEX (Romero-Garcés et al. 2015b), RoboCog (Romero-Garcés et al. 2015a)].
 - Tutoría [DAC (Vouloutsis et al. 2015)].
 - Evaluación médica [LIDA (Madl y Franklin 2015), RoboCog (Bandera et al. 2016)].
4. Arquitecturas de motivación biológica se centran en el aspecto evolutivo de las habilidades físicas y la reconstrucción sensoriomotora:
 - Plataforma robótica infantil iCub explora la adquisición de habilidades de locomoción, agarre y manipulación (Albus 1994)
 - Robots Dav y SAIL aprenden navegación guiada por visión [SASE (Weng y Zhang 2002)]
 - ISAC aprende affordances de agarre (Ulutas et al. 2008).

2. Arquitecturas cognitivas: Aplicaciones

Human-robot and human/computer interaction

- La Interacción humano-robot (Human-robot interaction, HRI) es un campo multidisciplinar que estudia diversos aspectos de la comunicación entre personas y robots.
- Muchas de estas interacciones se estudian en el contexto de la robótica social, asistencial o del desarrollo.
- Dependiendo del nivel de autonomía que demuestre el robot, las interacciones van desde el control directo (teleoperación) hasta la plena autonomía del robot, lo que permite la colaboración entre iguales.
- Arquitecturas cognitivas en el campo de la HRI, son fundamentales para desarrollar robots que puedan interactuar de manera efectiva con los humanos en una variedad de entornos y situaciones. **Procesar y entender las señales humanas, tomar decisiones y aprender de las interacciones.**

2. Arquitecturas cognitivas: Aplicaciones

Human-robot and human/computer interaction

- El robot debe de reconocer, comprender y participar en situaciones de comunicación:
 - **Explícita.** El humano se dirige verbalmente al robot.
 - **Implícita.** El humano señala un objeto.
- El robot debe ser capaz de participar en acciones conjuntas:
 - **Proactiva.** Planificando y proponiendo los planes resultantes al humano.
 - **Reactiva.** Respondiendo adecuadamente a las acciones e indicaciones humanas.
- Moverse y actuar de forma segura, eficiente y legible, teniendo en cuenta reglas sociales como la proxémica.

2. Arquitecturas cognitivas: Aplicaciones

Human-robot and human/computer interaction

Comunicación, acción conjunta y ejecución consciente del ser humano, componen la investigación sobre la interacción humano-robot. Pueden entenderse desde el punto de vista de las habilidades cognitivas que exigen.

La acción conjunta, por ejemplo, parte de:

- Un objetivo conjunto, previamente establecido y acordado (normalmente a través del diálogo).
- Un entorno físico, estimado a través de las capacidades de detección exteroceptivas del robot, y aumentado por inferencias extraídas de observaciones previas.
- Un estado de creencias que incluye conocimientos de sentido común a priori y modelos mentales de cada uno de los agentes implicados (el robot y sus compañeros humanos).

2. Arquitecturas cognitivas: Aplicaciones

Natural language processing (NLP)

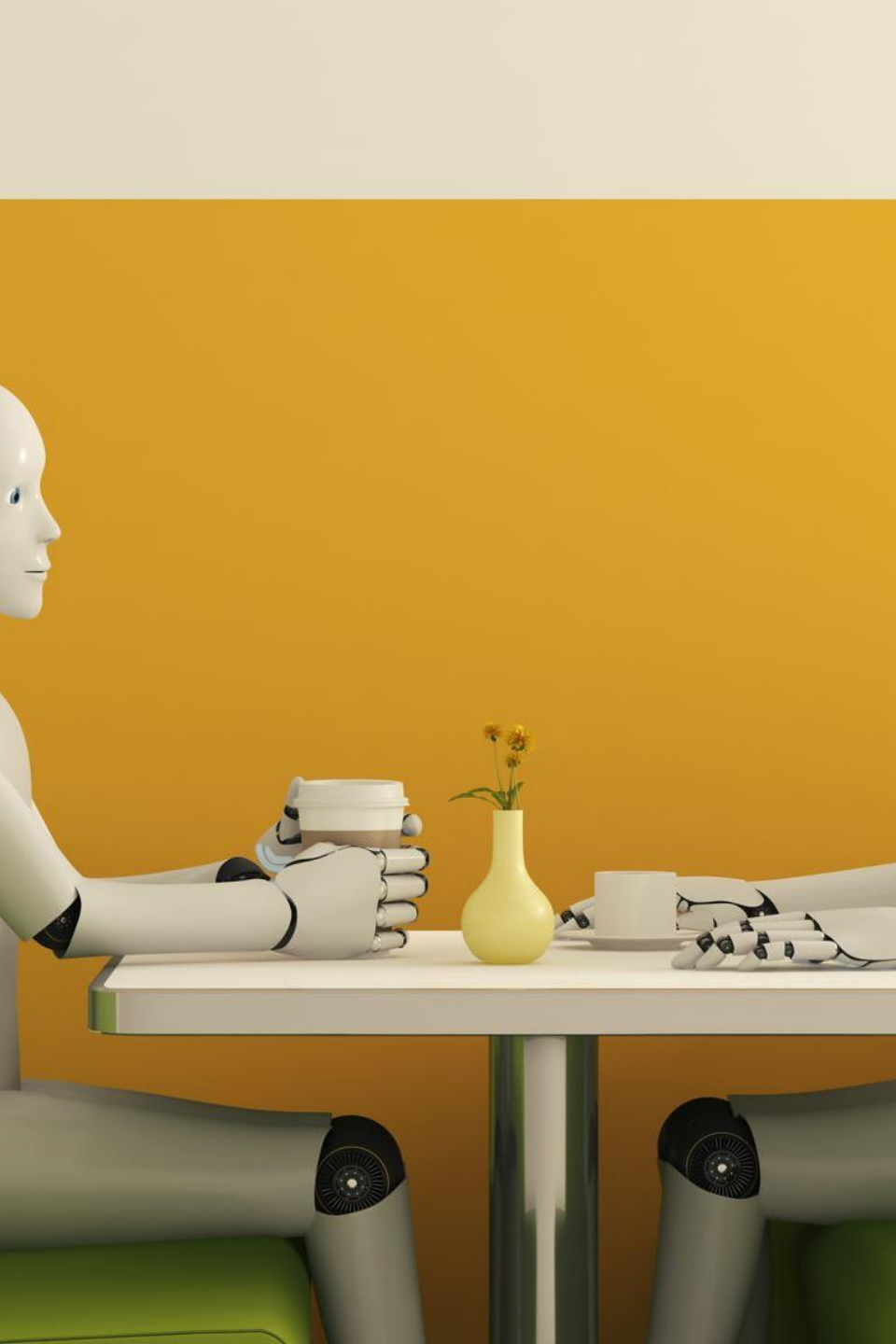
El procesamiento del lenguaje natural es un amplio campo multidisciplinar que **estudia la comprensión del lenguaje escrito o hablado**. En el contexto de las arquitecturas cognitivas, se han considerado muchos aspectos del NLP, desde la percepción auditiva de bajo nivel, el análisis sintáctico y la semántica hasta la conversación en dominios limitados.

- **Primeros modelos de traducción automática** (*Georgetown-IBM*, 1950).
- **Era de los sistemas basados en reglas (1960-1970)**. Grandes gramáticas escritas a mano (por lingüistas y programadores), redes semánticas y marcos (frames) para representar el conocimiento del mundo.
 - Problemas: Altísimo coste de construcción y mantenimiento de reglas. Dificultad para manejar ambigüedad, variabilidad y dominio abierto.
 - Ejemplos:
 - **ELIZA (Weizenbaum, 1966)**: imitaba el estilo de un terapeuta centrado en el paciente, donde ante cada entrada del usuario, ELIZA aplicaba transformaciones de texto y devolvía preguntas o reformulaciones que invitaban al interlocutor a profundizar en sus propios pensamientos.

2. Arquitecturas cognitivas: Aplicaciones

Natural language processing (NLP)

- **Transición a métodos estadísticos y de aprendizaje (años 1980–2000)**
 - Modelos de n-gramas y corpus etiquetados
 - Introducción de técnicas de aprendizaje supervisado
- **Revolución del Deep Learning y LLMs (desde 2018)**
 - Modelos preentrenados de gran escala
 - Transformer (Vaswani et al., 2017) como base de:
 - **GPT (OpenAI)**: generación autoregresiva de texto.
 - **BERT (Google)**: codificación bidireccional para comprensión.
 - LLaMA (Meta), PaLM (Google), ChatGPT, etc., con miles de millones de parámetros.
 - Desafíos:
 - Explicabilidad de las decisiones.
 - Alucinaciones.
 - Sesgos en los datos de entrenamiento.



3. Interacción humano-máquina

3. Interacción humano-máquina

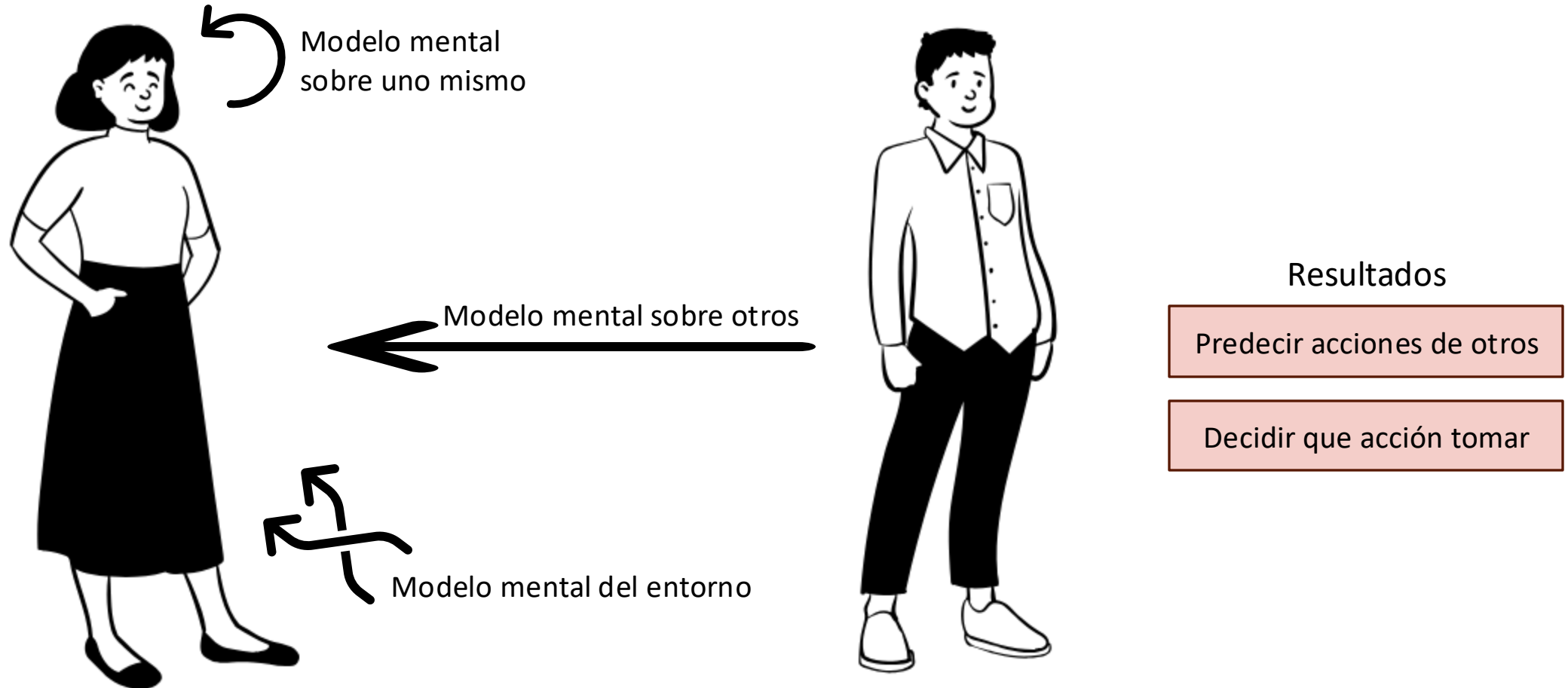
La cognición que hemos visto hasta ahora:

- Centrada en el agente
- Actividad "solitaria"
- Explorar e interactuar con el mundo: adaptación y autonomía

Actividad social no colectiva



3. Interacción humano-máquina



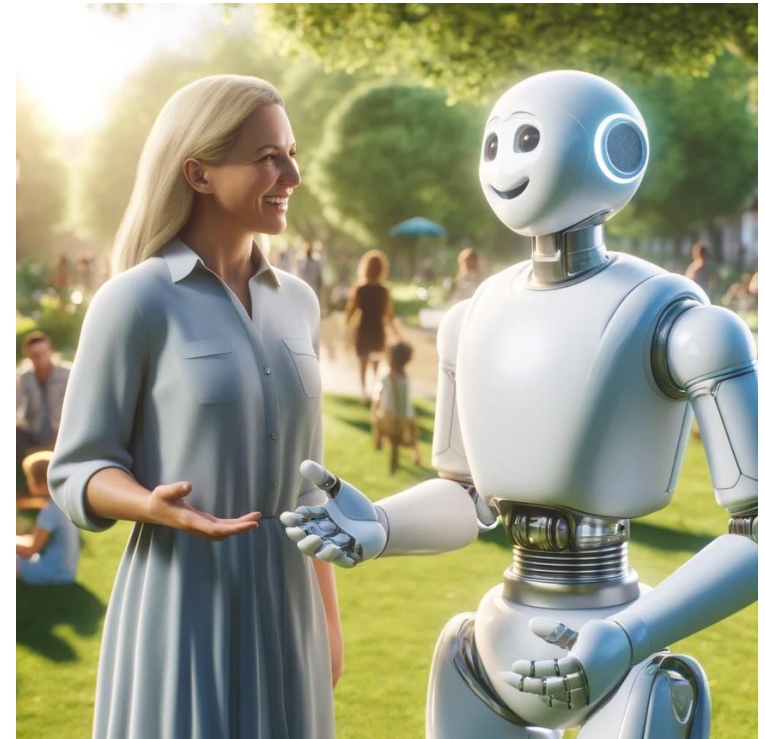
3. Interacción humano-máquina

Modelo del robot

Tipo de modelo	Contenido principal	Ejemplo de uso
Modelo del usuario	Preferencias, habilidades, intenciones	Ajustar velocidad de trabajo según experiencia del operario.
Modelo del entorno	Mapa 3D, objetos, zonas de seguridad	Planificar trayectorias que eviten colisiones.
Modelo de sí mismo	Estado de baterías, fuerzas máximas, precisión de sensores	Decidir cuándo recargar o recalibrar.

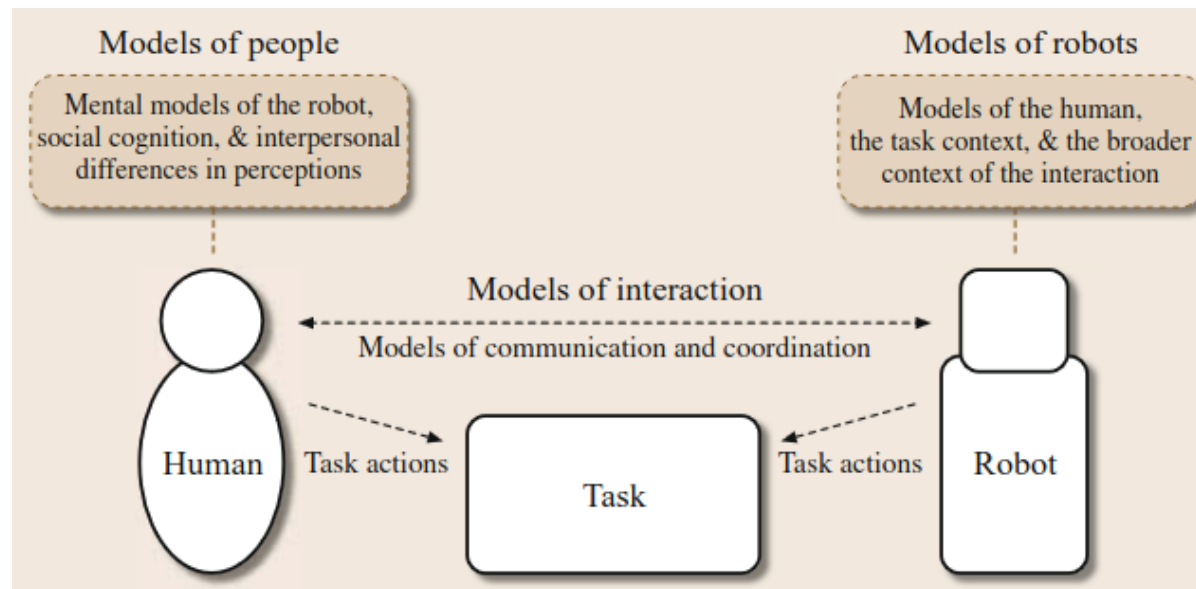
3. Interacción humano-máquina

La **interacción cognitiva humano-robot (HRI)** es un campo de investigación que pretende mejorar las interacciones entre los robots y sus usuarios mediante el desarrollo de modelos cognitivos para robots, que permitan comprender y procesar información de manera similar a como lo hacen los humanos.



3. Interacción humano-máquina

Modelos mentales del robot



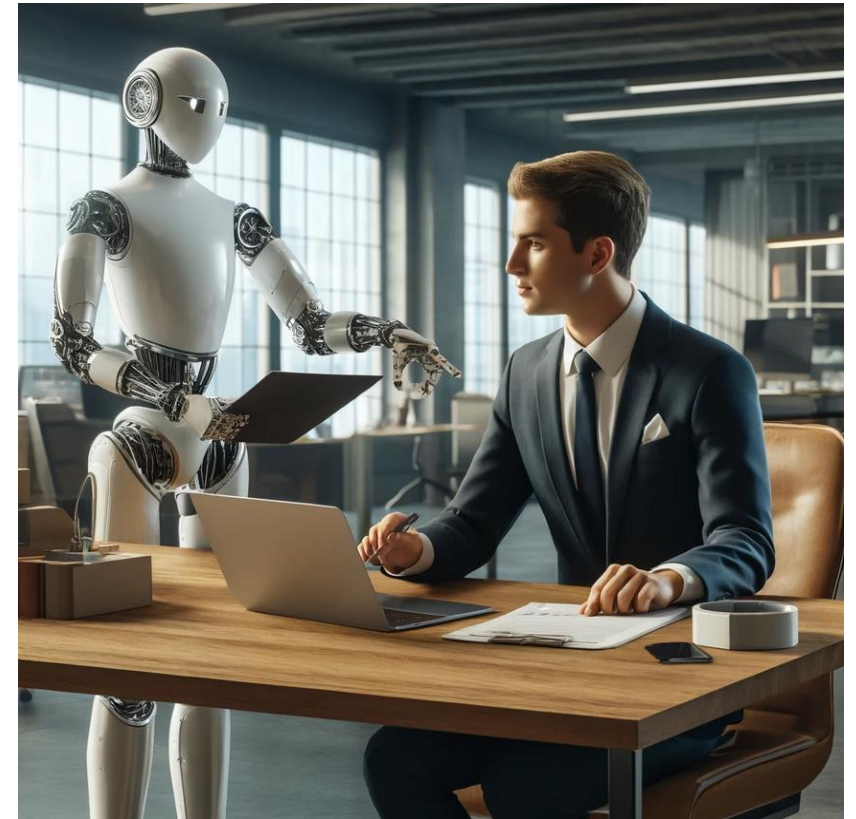
Handbook of Robotics (2008)

Actividades de investigación en HRI cognitivo

3. Interacción humano-máquina

Modelos mentales del robot

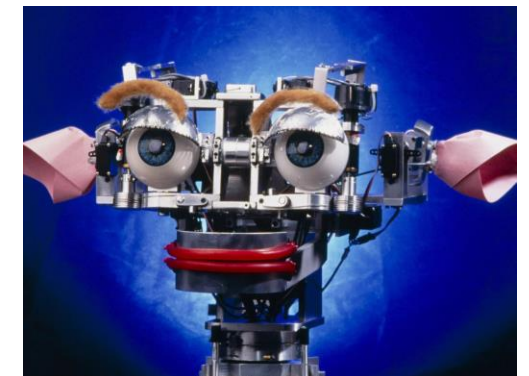
- Los comportamientos de las personas hacia la tecnología a menudo siguen las reglas sociales establecidas en las interacciones entre humanos.
- Las personas probablemente interpretarán los comportamientos interactivos de los robots de manera social.
- Es necesario comprender la cognición social humana para desarrollar plataformas robóticas que se adapten a las expectativas y comportamientos de los usuarios.
- Entender qué modelos mentales usan las personas para interpretar el comportamiento de los robots.



3. Interacción humano-máquina

Modelos atribuidos al robot

- En los trabajos de Sherry Turkle (MIT), se estudiaron el comportamiento de personas (niños y ancianos) hacia robots sociales como Kismet, Cog, PARO, Furby o My Real Baby.



3. Interacción humano-máquina

Modelos atribuidos al robot

Se producen dos principales aproximaciones:

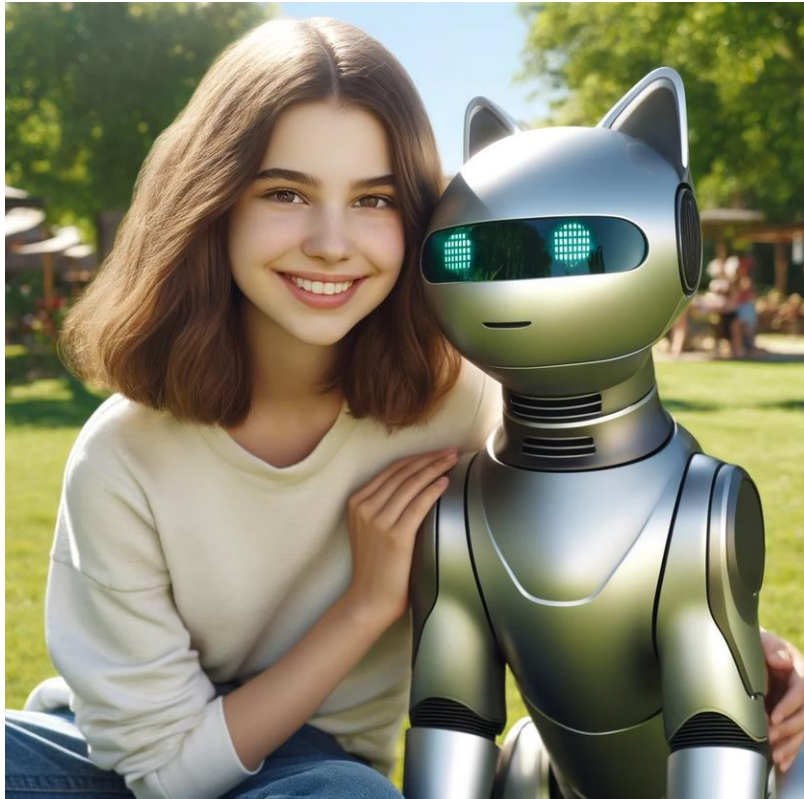
1. Enfoque “científico-exploratorio” (mechanistic stance)

- La persona interpreta al robot como un objeto técnico, sin emociones ni vida interior.
- Se centra en sus partes, su programación y sus causas mecánicas (por qué gira la cabeza, qué sensor ha activado), manteniendo “distancia emocional”.

2. Enfoque “relacional-animista” (relational-animistic approach)

- La persona atribuye al robot de cualidades de ser vivo: emociones, intenciones y necesidades. Se **antropomorfiza** al robot de forma deliberada.
- Trata al robot como un bebé o una mascota, respondiendo de forma afectiva (acallar un llanto de My Real Baby, acariciar a PARO, etc.), incluso cuando verbalmente afirma “sé que no es un ser vivo”.

3. Interacción humano-máquina



La **antropomorfización** es el proceso mediante el cual atribuimos características humanas a objetos no humanos (animales, objetos inanimados o, en este caso, robots). En el contexto de la interacción entre humanos y robots, la antropomorfización puede ocurrir de varias formas y tener diferentes efectos.

- **Diseño físico**
- **Comportamiento**
- **Interacción lingüística.**

3. Interacción humano-máquina

La **antropomorfización** en la interacción humano-robot puede tener varios **efectos**:

- **Mayor aceptación:** Cuando los robots parecen y se comportan de manera más humana, es más probable que las personas los acepten y se sientan cómodas interactuando con ellos.
- **Empatía:** La antropomorfización puede llevar a las personas a sentir empatía hacia los robots, especialmente cuando los robots están diseñados para parecer vulnerables o expresar emociones.
- **Expectativas exageradas:** Sin embargo, también existe el riesgo de que las personas atribuyan cualidades humanas a los robots que no poseen, lo que puede llevar a expectativas exageradas sobre las capacidades del robot o a malentendidos sobre la naturaleza de la interacción.
- **Mejora de la comunicación:** Al imitar el comportamiento humano, los robots pueden facilitar la comunicación con las personas y hacer que la interacción sea más natural y efectiva.

3. Interacción humano-máquina

¿Qué características favorecen la antropomorfización?

1. Embodiment físico

- Un robot físico despierta más proyección de vida e intenciones que un agente puramente virtual.

2. Co-presencia y entorno compartido

- Cuando el robot está “a mi lado” o comparte mi mismo espacio (oficina, salón, taller), siento que sus acciones tienen relevancia social directa.

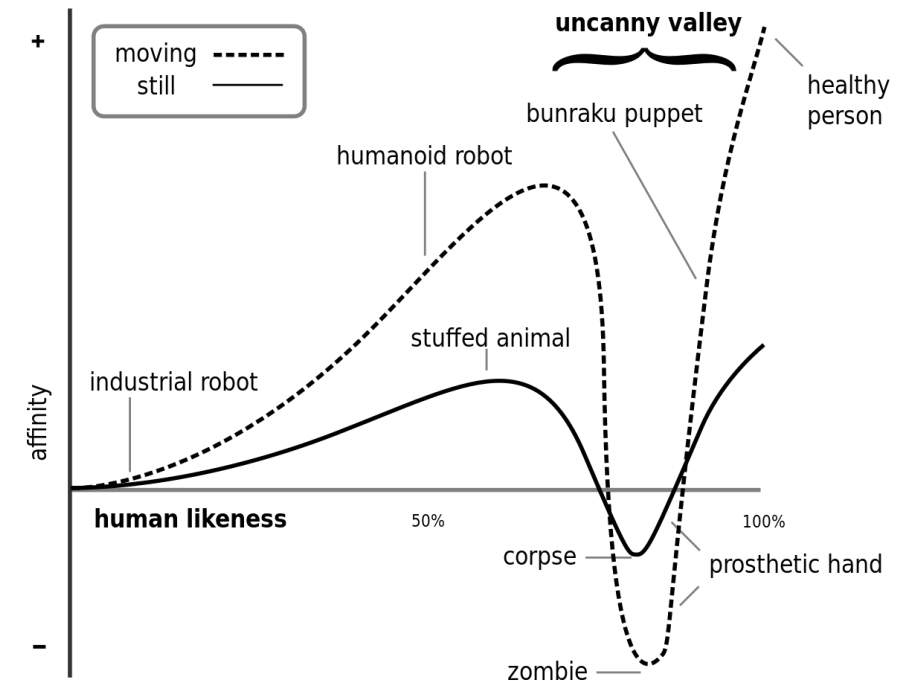
3. Normas de cortesía y etiqueta

- Robots que saludan, esperan turno para hablar o respetan el espacio personal refuerzan la idea de “compañero social” y no solo “herramienta”.

3. Interacción humano-máquina

Uncanny valley (el valle inquietante)

- El valle inquietante (*Masahiro Mori, 1970*) es una teoría que sugiere una conexión entre la apariencia de un objeto y la reacción emocional hacia éste, basado en su similitud con un ser humano.
- Cuando los robots humanoides se parecen demasiado en aspecto y comportamiento a un ser humano real, generan una sensación de repulsión en las personas que los observan.



3. Interacción humano-máquina

Modelos de interacción

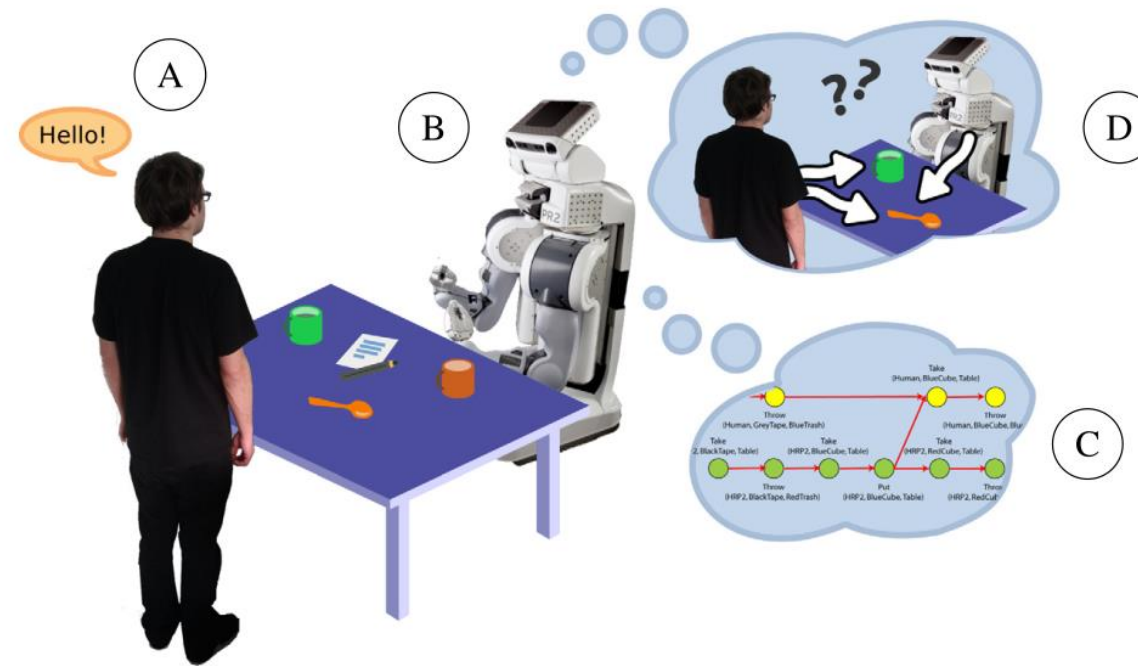
- Los robots humanoides son diseñados para una interacción efectiva con seres humanos.
- Los robots deben contar con modelos que definan características físicas y cognitivas de su entorno, para alcanzar sus metas.
- Existen diferentes categorías de modelos de interacción humano-robot. Entre ellos nos encontramos con los siguientes:
 1. Basados en diálogos
 2. Para planificar acciones

3. Interacción humano-máquina

Basados en diálogos

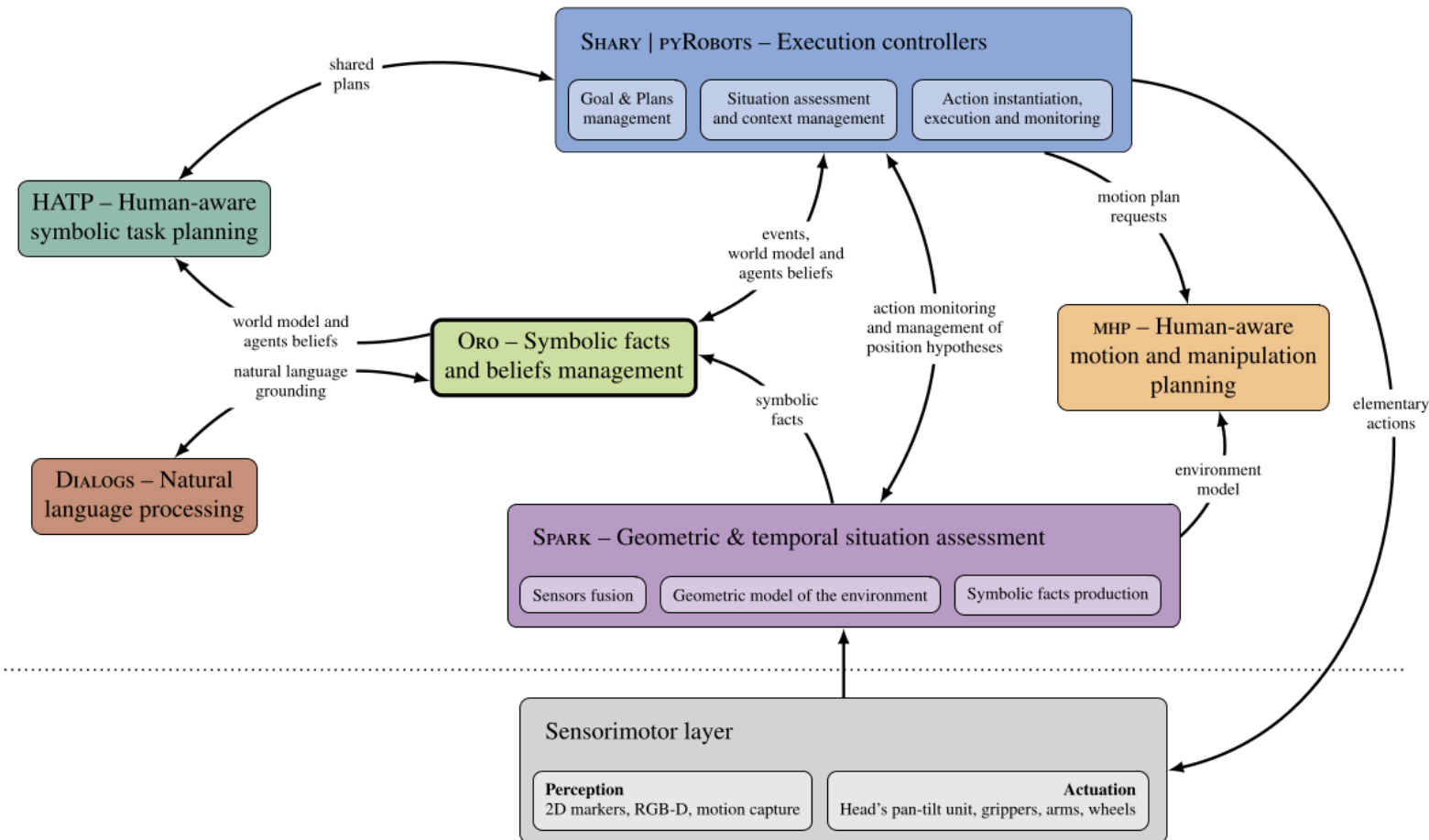
- Establecer un modelo común en actividades conjuntas entre humanos y robots. Estos modelos buscan facilitar la comprensión mutua y mejorar la colaboración en tareas específicas.
- Tareas como **la navegación y la exploración colaborativa**, utilizando el diálogo para compartir información y tomar decisiones conjuntas sobre el control en puntos clave de la tarea.
- Abordar la tarea del dominio y del diálogo como acción conjunta. Estos modelos integran **módulos de interpretación semántica y toma de decisiones centralizada**, utilizando recursos como la historia del discurso en curso, un modelo del mundo y un plan de dominio para generar comportamientos de acción y comunicación. (Ejemplo, Foster et al. (2006).

3. Interacción humano-máquina



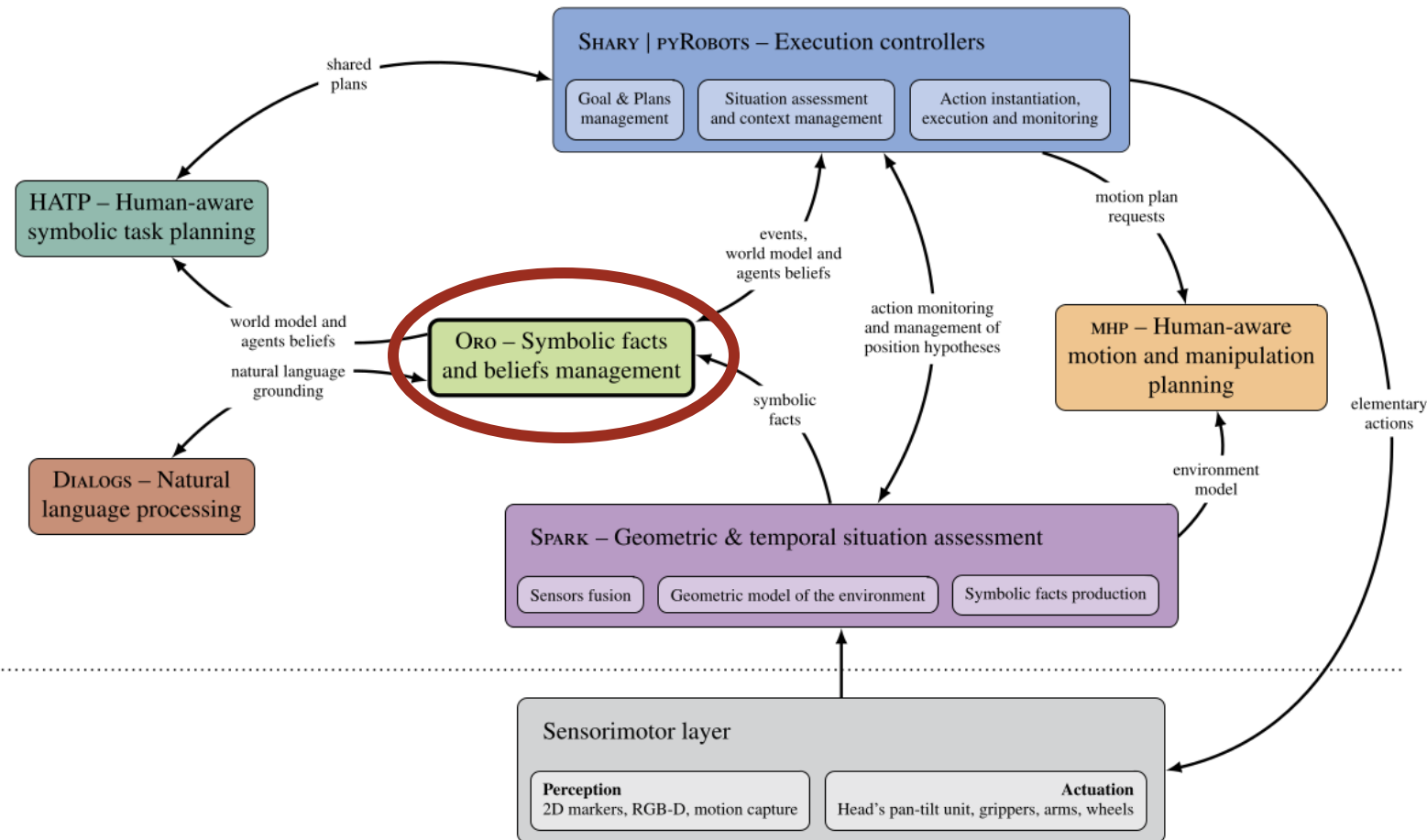
S. Lemaignan et al. / Artificial cognition for social human–robot interaction: An implementation / Artificial Intelligence 247 (2017) 45–69

3. Interacción humano-máquina



S. Lemaignan et al. /
Artificial cognition for
social human–robot
interaction: An
implementation /
Artificial Intelligence
247 (2017) 45–69

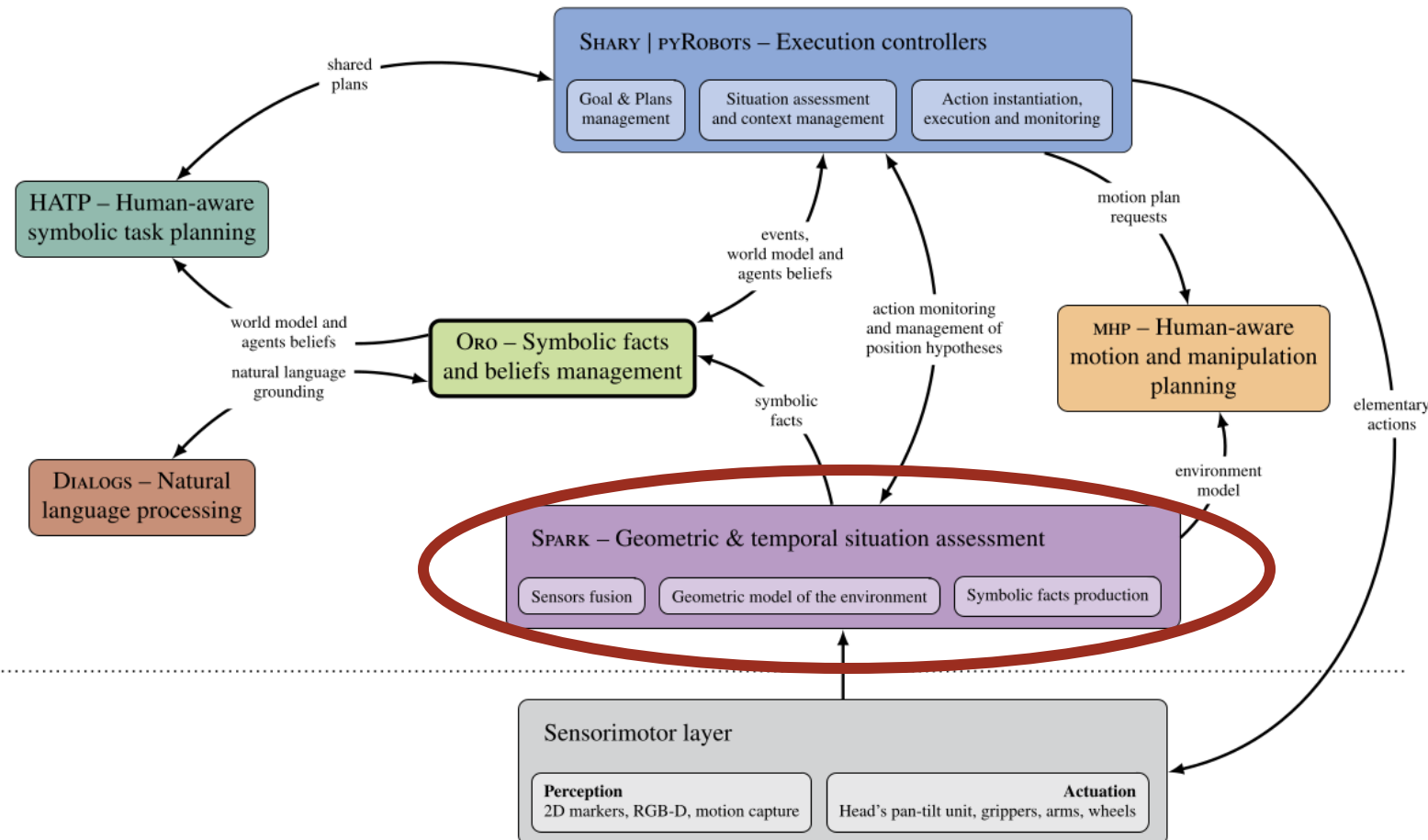
3. Interacción humano-máquina



ORO (Base de Conocimiento):

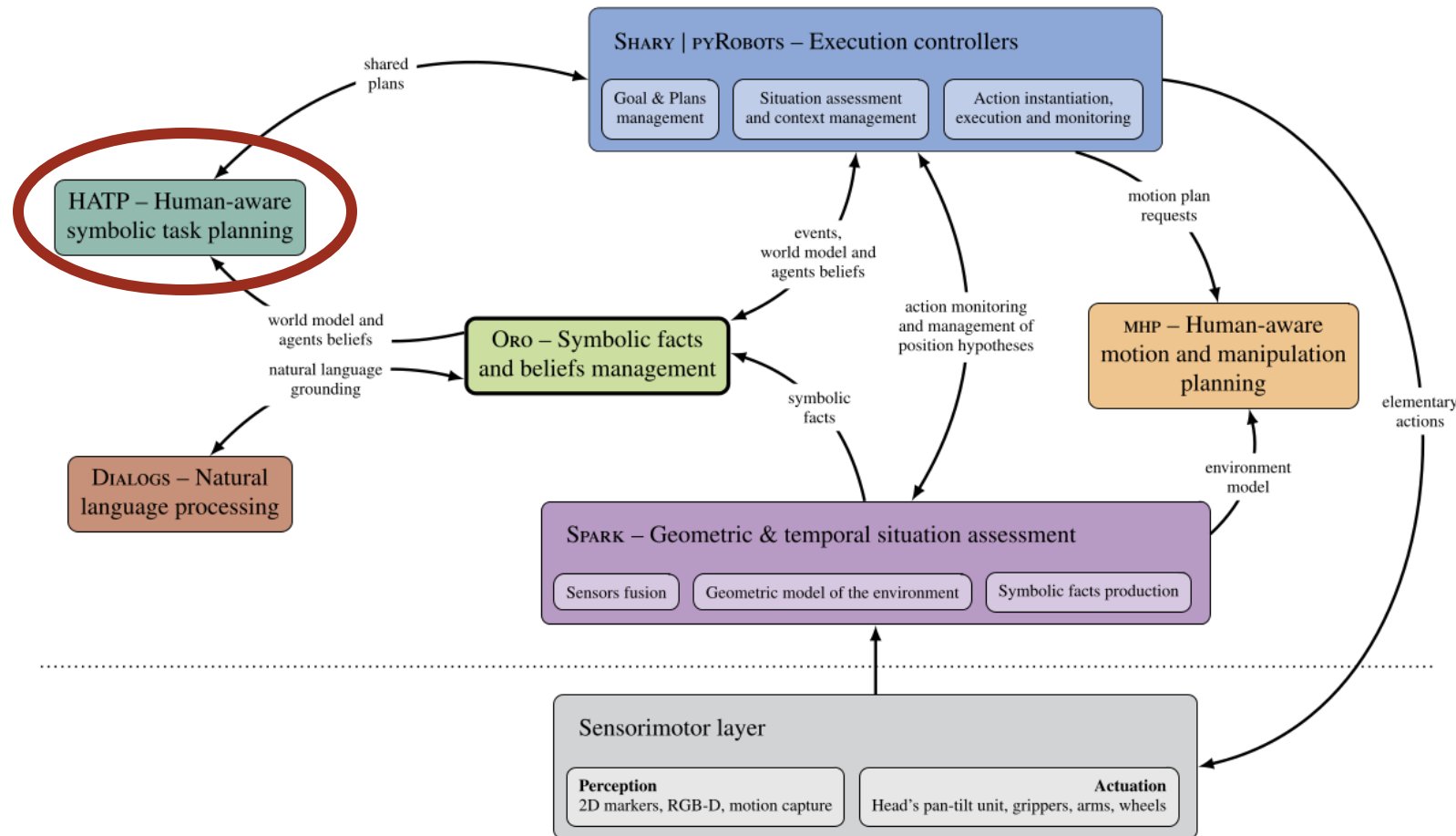
- Almacena el **conocimiento simbólico**.
- Maneja y procesa el conocimiento de forma simbólica y lógica

3. Interacción humano-máquina



SPARK. Razonamiento espacial y la evaluación del entorno. Realiza evaluación de la situación física en el entorno del robot, lo que le permite generar conocimiento simbólico sobre la relación espacial entre los objetos y los agentes (humanos y robots).

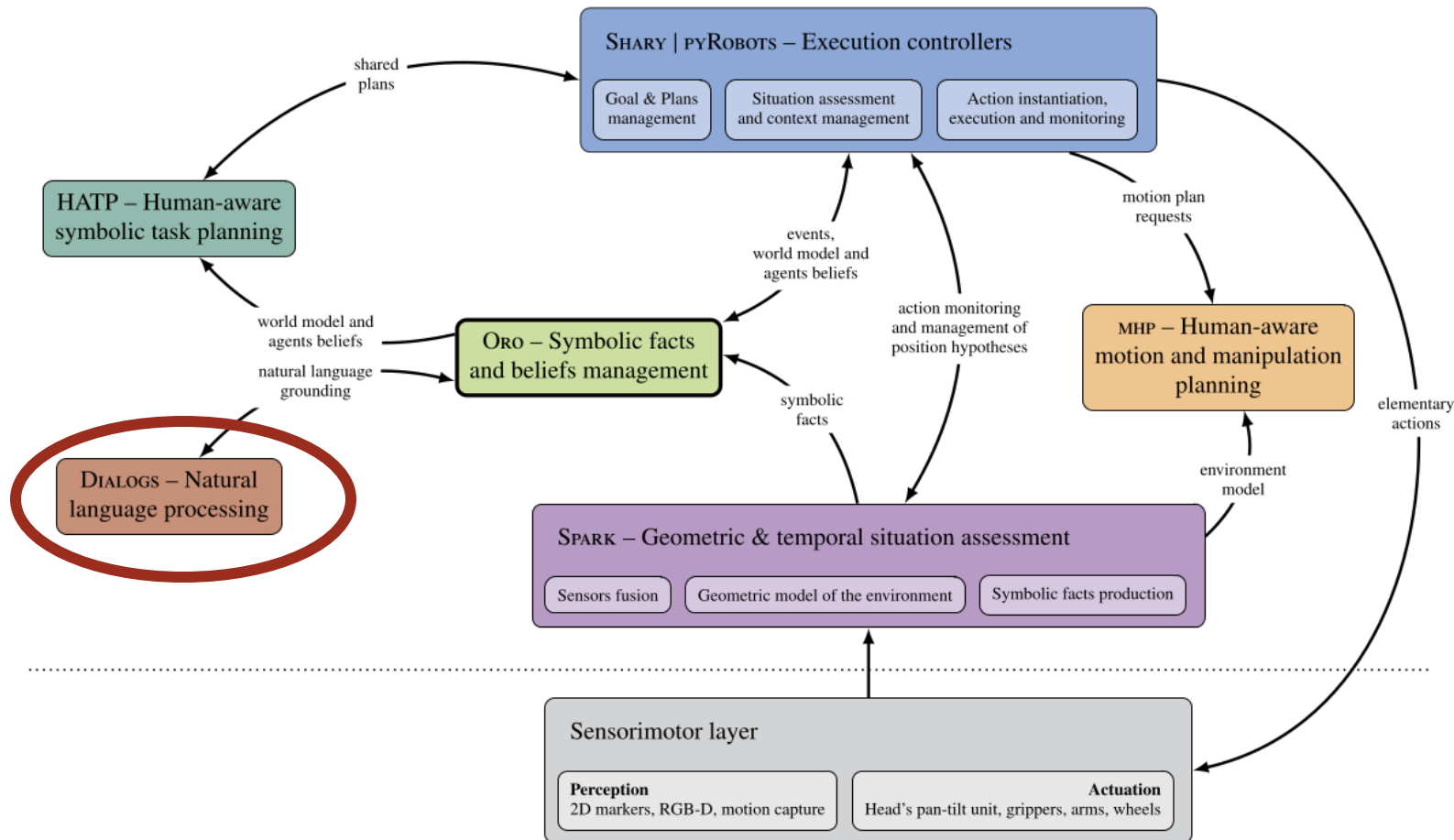
3. Interacción humano-máquina



HATP (Planificador Simbólico de Tareas)

- Decide qué **acciones debe realizar** el robot para alcanzar un objetivo dado, basándose en un conjunto de reglas y hechos.
- Produce **planes colaborativos** (sincronización)
- **Asegura que las tareas** cumplan las condiciones de éxito de las acciones planeadas.

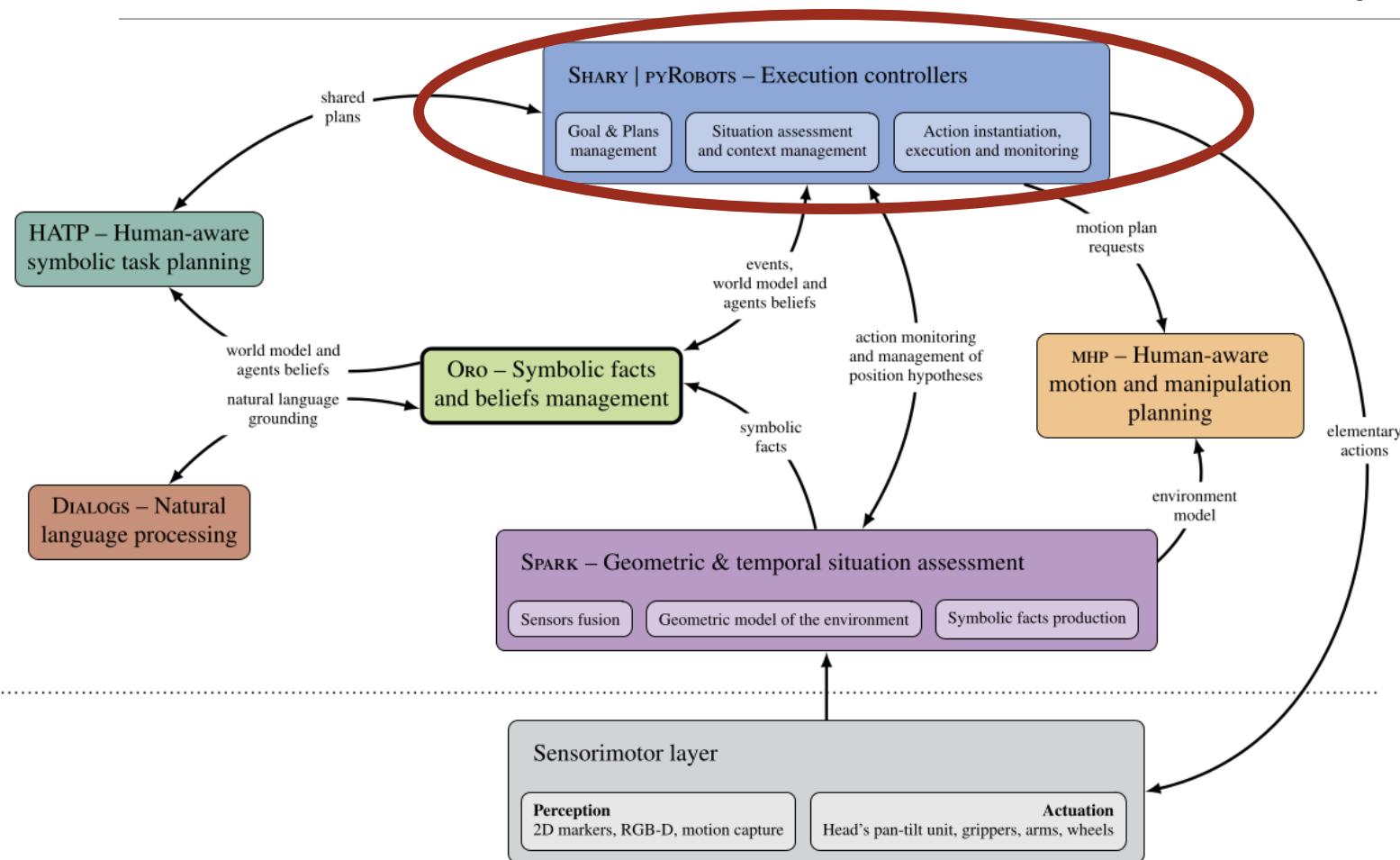
3. Interacción humano-máquina



DIALOGS (Procesamiento de Diálogo y Comunicación Multimodal)

- **Procesamiento de lenguaje natural** y la interacción verbal entre el robot y los humanos.
- **Resuelve ambigüedades**, entiende las intenciones del humano, y usa el conocimiento en la base de datos para interpretar y generar respuestas apropiadas.

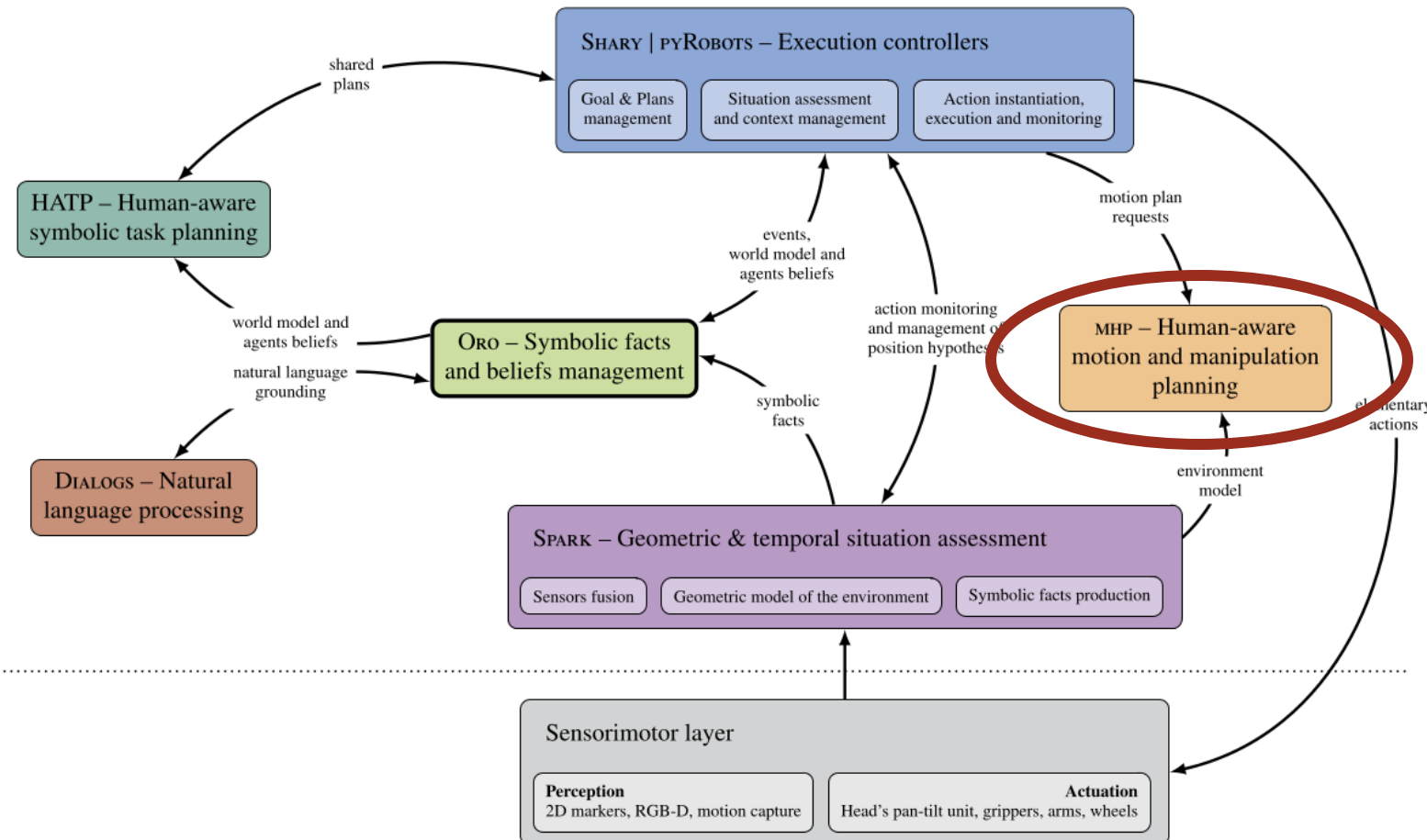
3. Interacción humano-máquina



SHARY. Controlador de ejecución. Su principal función es **gestionar la ejecución de planes y coordinación de acciones**.

- **Ejecutar** las acciones físicas del robot y supervisar su progreso.
- Controla la **ejecución de planes colaborativos** elaborados por HATP, asegurando que se coordinen las acciones entre el robot y los humanos

3. Interacción humano-máquina



MHP (Human-Aware Motion Planning)

- Complementa el control de ejecución en la arquitectura del robot.
- **Planificar los movimientos geométricos del robot** de manera que no solo sean eficientes y seguros, sino también **socialmente aceptables** para los humanos con los que interactúa.

3. Interacción humano-máquina

Acción conjunta en HRI

Meta común

- Establecida y acordada previamente (normalmente vía diálogo)

Entorno físico

- Percibido por los sensores exteroceptivos del robot
- Enriquecido con inferencias de observaciones previas

Estado de creencias

- Conocimiento a priori (sentido común)
- Modelos mentales compartidos de todos los agentes (robot y humano)

3. Interacción humano-máquina

Basados en diálogos

Jido, un manipulador móvil monobrazo, observa mientras dos personas transportan cajas y responde a preguntas sobre la ubicación de objetos específicos.

- **Ejemplo1.** Desambiguación implícita a través de perspectiva visual.

MAN A. "Jido, can you give me the video tape?"

`find(?obj type VideoTape) ?obj = [BLACK_TAPE, WHITE_TAPE]`

Aunque existe ambigüedad desde la perspectiva del robot, el humano utiliza el cuantificador definido "the" al referirse a una cinta de video, lo que el procesador de lenguaje natural interpreta como una referencia a un objeto conocido, es decir, al único visible en el modelo de conocimiento humano.



Artificial cognition for social human-robot interaction: An implementation (S.Lemaignan et al)

3. Interacción humano-máquina

Basados en diálogos

- **Ejemplo2.** Desambiguación explícita a través de interacción verbal y gestos.

Man B: "Jido, what's in the box?"

$\text{find}(\text{?obj type Box}) \Rightarrow \text{?obj} = [\text{CARDBOARD_BOX}, \text{TOOLBOX}]$

Jido: "Which box, the toolbox or the cardboard box?"

Man B: "This box" (Pointing at it)

$\text{pointsAt}(\text{?ag}, \text{?obj}) \wedge \text{looksAt}(\text{?ag}, \text{?obj}) \rightarrow \text{focusesOn}(\text{?ag}, \text{?obj})$

$\text{find}(\text{?obj type Box}; \text{USER_B focusesOn ?obj}) \Rightarrow \text{?obj} = [\text{CARDBOARD_BOX}]$

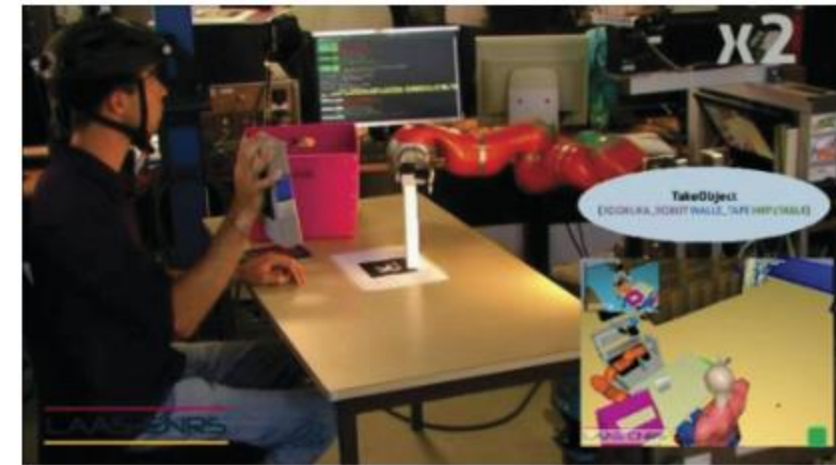


Artificial cognition for social human-robot interaction: An implementation (S.Lemaignan et al)

3. Interacción humano-máquina

Planificación de acciones

- **Ejemplo.** La tarea consiste en limpiar una mesa de forma cooperativa moviendo los objetos a sus ubicaciones de destino.
- planificador simbólico de tareas + controlador de ejecución => ejecución de un plan compartido
- Una sola cinta de vídeo sobre una mesa. La cinta de vídeo sólo es accesible para el robot, mientras que la papelera (a la que se supone que deben trasladarse los objetos) sólo es accesible para el ser humano: el robot debe elaborar un plan compartido que implique una acción conjunta.

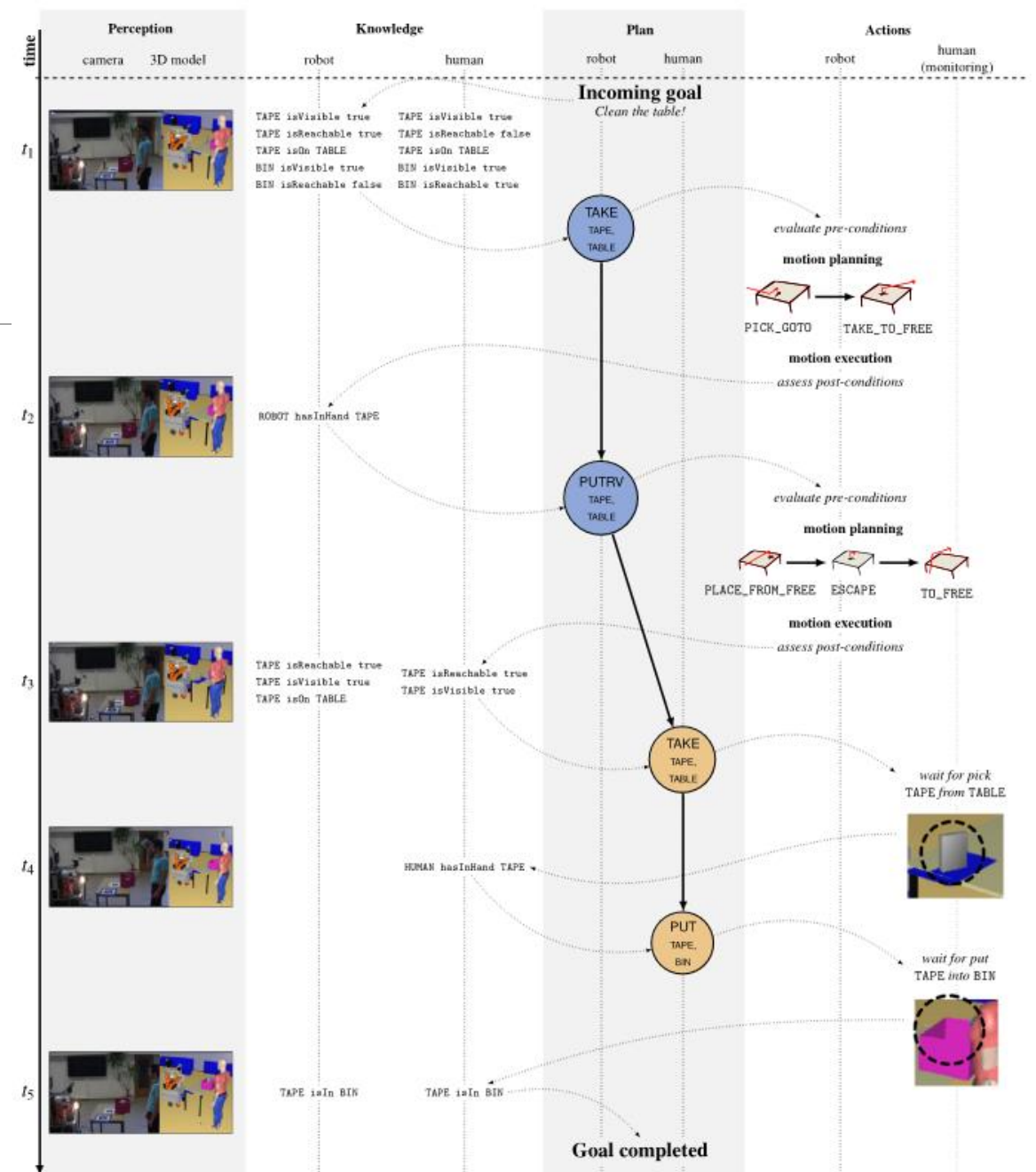


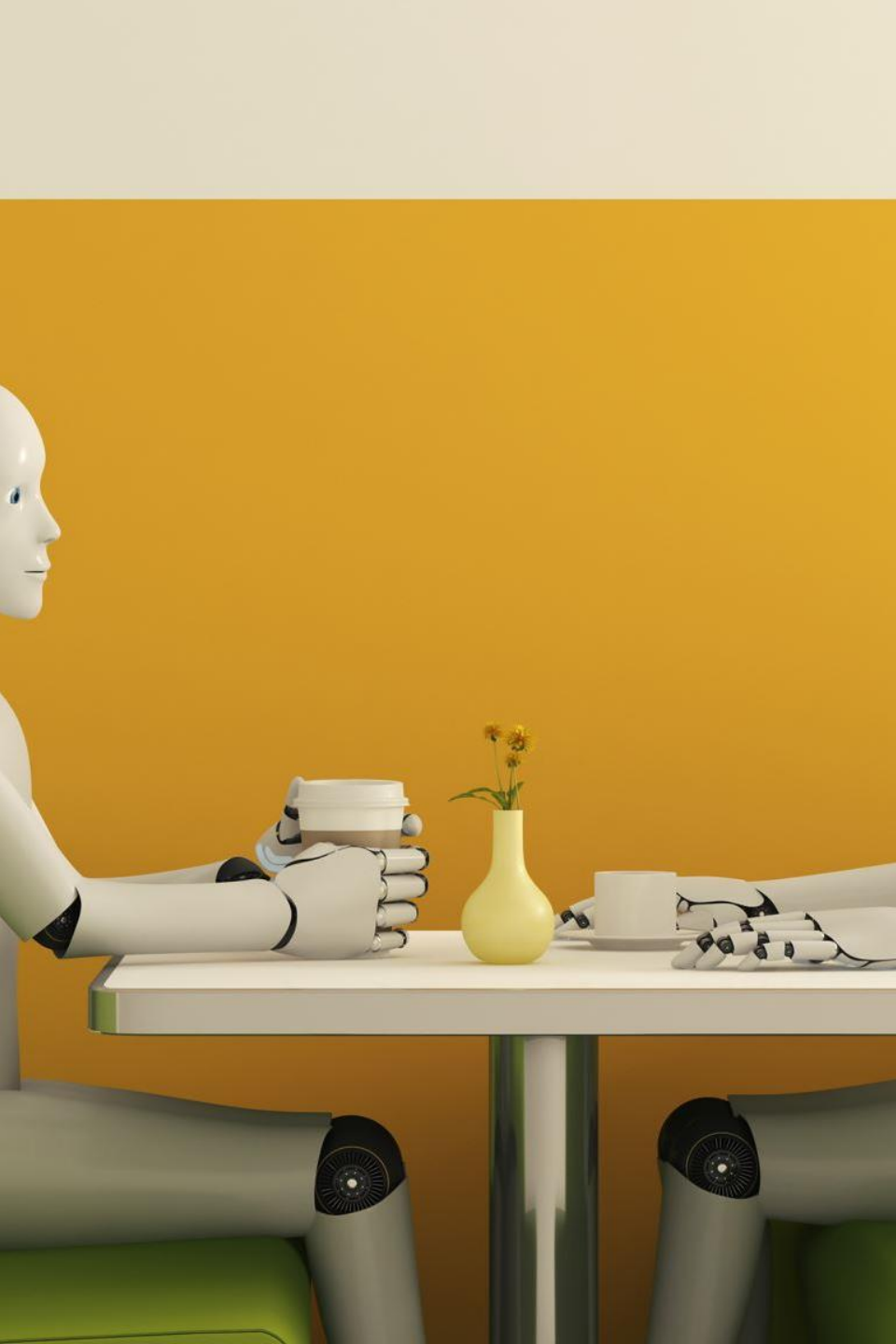
Artificial cognition for social human-robot interaction: An implementation (S.Lemaignan et al)

3. Interacción humano-máquina

Planificación de acciones

Artificial cognition for social human-robot interaction: An implementation (S.Lemaignan et al)





Referencias bibliográficas

Referencias bibliográficas

- Séverin Lemaignan, Mathieu Warnier, E. Akin Sisbot, Aurélie Clodic, Rachid Alami. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, 247, pp 45-69 (2017).
- Vernon, D., "Cognitive Architectures" in *Cognitive Robotics*, A. Cangelosi and M. Asada (Eds.), MIT Press, in press.
- Kotseruba, I., Tsotsos, J.K. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artif Intell Rev* **53**, 17–94 (2020)
- M.E. Foster, T. By, M. Rickert, A. Knoll: Human-robot dialogue for joint construction tasks, *Proc. 8th Int. Conf. Multimodal Interfaces*, Banff (2006) pp. 68–71
- Lemaignan, Séverin et al. "Artificial cognition for social human-robot interaction: An implementation." *Artif. Intell.* 247 (2017): 45-69.