

Ingeniero en Software y Tecnologías Emergentes

Inteligencia Artificial

Unidad 2 Agentes Inteligentes

ISyTE

Dr. Héctor Zatarain Aceves

hector.zatarain@uabc.edu.mx

Motivación

“La Inteligencia Artificial Generativa es solo el comienzo, los agentes de IA es lo que viene a continuación”

aoud Abdel.

-D

“Los agentes de IA se convertirán en nuestros asistentes digitales, ayudándonos a navegar por las complejidades del mundo moderno. Harán nuestras vidas más fáciles y eficientes.”

Jeff Bezos, Fundador y CEO de Amazon.

-J

“Los agentes no solo cambiarán la forma en que todos interactuamos con las computadoras. También revolucionarán la industria del software, provocando la mayor revolución informática desde que pasamos de escribir comandos a tocar iconos. Los agentes no solo harán recomendaciones; te ayudarán a ponerlas en práctica.”

-Bill Gates.

“La inteligencia artificial de agentes resolverá de forma autónoma el 80% de los problemas comunes de servicio al cliente sin intervención humana para 2029.”

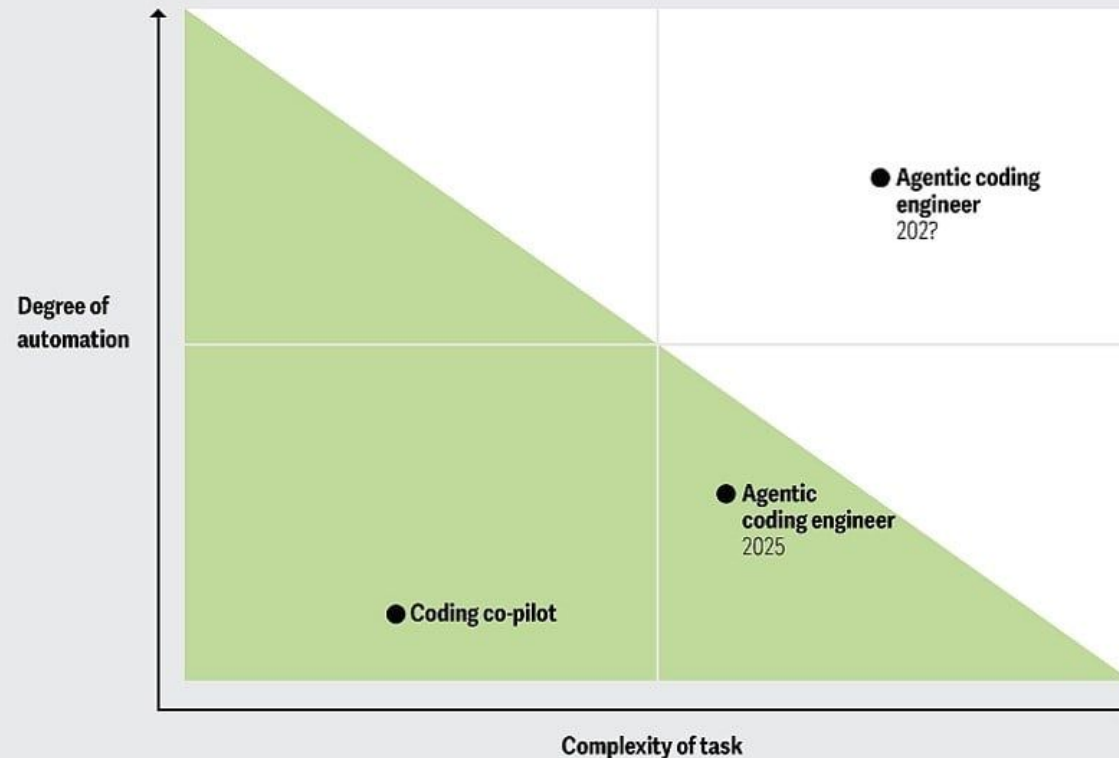
– **Gartner, [Comunicado de prensa Marzo 2025](https://www.gartner.com/en/newsroom/press-releases/2025-03-05-gartner-predicts-agentic-ai-will-autonomously-resolve-80-percent-of-common-customer-service-issues-without-human-intervention-by-20290)**

Motivación

Figure 1

Gen AI agents are advancing

● More human supervision required ● Less human supervision required



Source: Deloitte analysis.

Deloitte Insights | deloitte.com/insights

Los **agentes de IA generativos autónomos**, conocidos como “**agentes de IA**”, son soluciones de software que pueden completar tareas complejas y alcanzar objetivos con poca o ninguna supervisión humana.

Actualmente estos cometen demasiados **errores** para manejar trabajos completos, o incluso parciales, sin supervisión humana

Las grandes empresas tecnológicas y las nuevas empresas se esfuerzan por hacer que **agentes de IA** en su rol de ingenieros de software sean más autónomos y confiables, para que los codificadores humanos (y sus empleadores) puedan confiar en ellos para manejar partes de su carga de trabajo.

COMPETENCIA

Identificar la aplicación del enfoque de diseño orientado a agentes, para implementar un **sistema inteligente**, mediante **agentes de software** como unidades de abstracción principales, con honestidad, actitud creativa y propositiva.

Contenido

CONCLUSIÓN

- 2.1. Definición de Agentes Inteligentes.
- 2.2. Diferencias entre el Enfoque Orientado a Objetos y Orientado a Agentes
- 2.3. Estructura y Comportamiento de los Agentes Inteligentes.
- 2.4. Plataformas para desarrollar Agentes Inteligentes
- 2.5. Aplicaciones de sistemas de Agentes Inteligentes

Agente

Un agente es cualquier entidad capaz de **percibir** su **medioambiente** con la ayuda de **sensores** y actuar en ese medio utilizando **actuadores**.

El término **percepción** se utiliza en este contexto para indicar que el agente puede recibir entradas en cualquier instante.

La **secuencia de percepciones** de un agente refleja el historial completo de lo que el agente ha recibido.

En general, un agente tomará una **decisión** en un momento dado dependiendo de la secuencia completa de percepciones hasta ese instante.

Agente

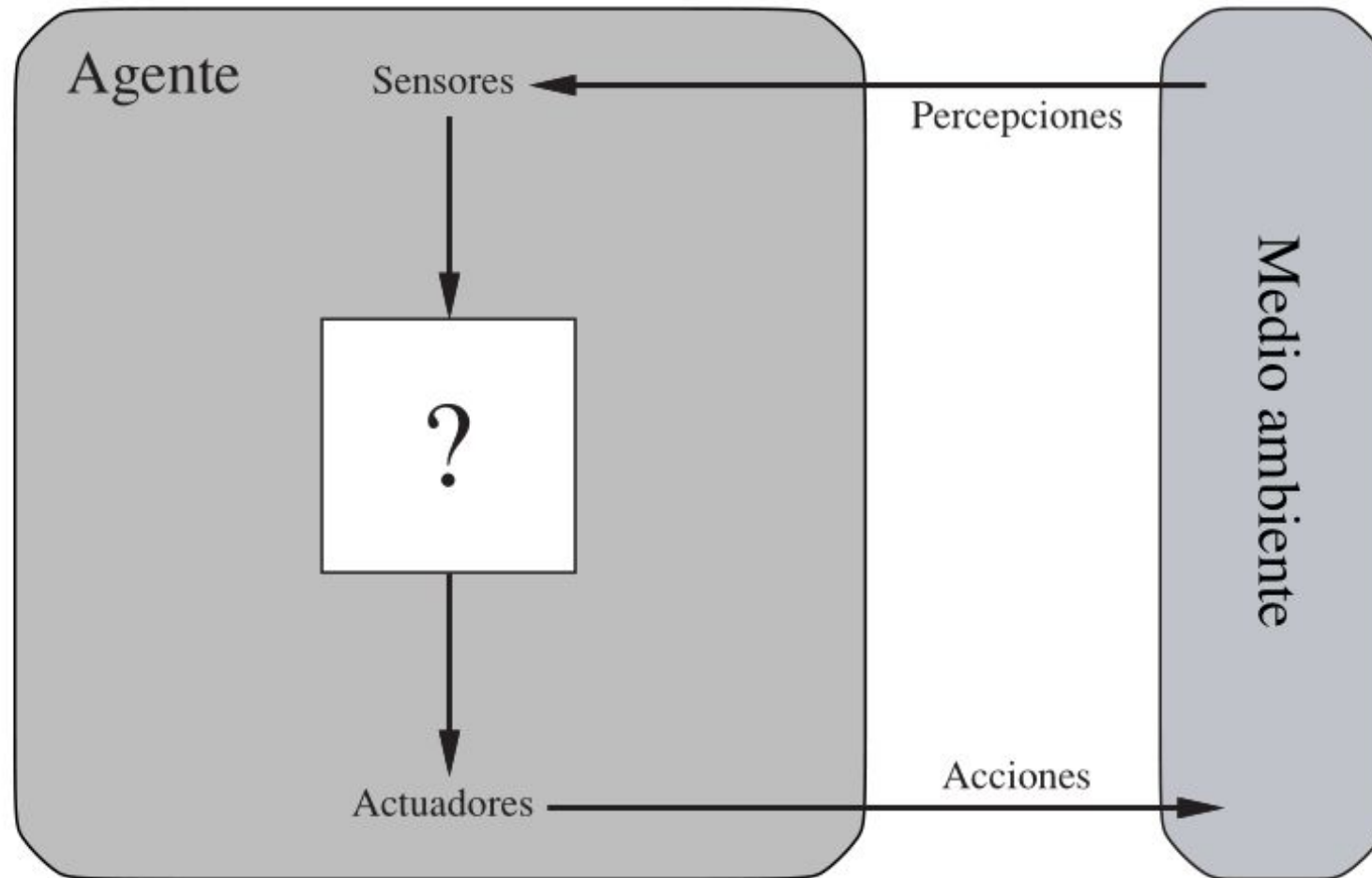


Figura 1. Los agentes interactúan con el medioambiente mediante sensores y efectores.

Agente

En términos matemáticos se puede decir que el comportamiento del agente viene dado por la **función** del agente que proyecta una percepción dada en una acción.

La función que describe el comportamiento de un agente se puede presentar en forma de **tabla**; en la mayoría de los casos esta tabla sería muy grande.

La tabla es una caracterización externa del agente. Inicialmente, la función del agente para un agente artificial se implementará mediante el **programa del agente**.

Es importante **diferenciar** estas dos ideas.

Agente

La **función del agente** es una descripción matemática abstracta.

El **programa del agente** es una implementación completa, que se ejecuta sobre la arquitectura del agente.

Ejemplo 1:

Supongamos una aspiradora autónoma. El mundo de la aspiradora cuenta con dos localizaciones solamente.

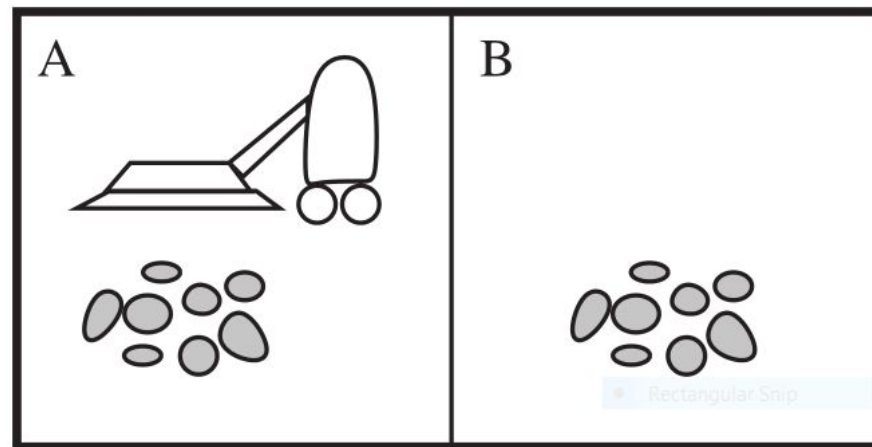
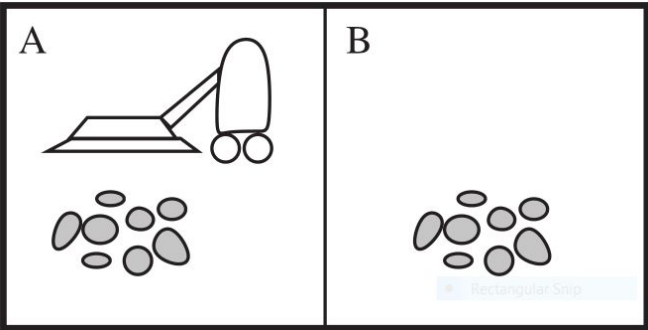


Figura 2. El mundo de la aspiradora con dos localizaciones solamente.

Agente



Secuencia de percepciones	Acción
[A, Limpio]	Derecha
[A, Sucio]	Aspirar
[B, Limpio]	Izquierda
[B, Sucio]	Aspirar
[A, Limpio], [A, Limpio]	Derecha
[A, Limpio], [A, Sucio]	Aspirar
—	—
—	—
—	—
[A, Limpio], [A, Limpio], [A, Limpio]	Derecha
[A, Limpio], [A, Limpio], [A, Sucio]	Aspirar
—	Rectangular Snip —
—	—
—	—

Figura 3. Tabla parcial de una función de agente sencilla para el mundo de la aspiradora que se muestra en la Figura 2.

Agente

Es posible definir **varios agentes** para el mundo de la aspiradora simplemente rellenando la columna de la derecha de formas distintas.

La pregunta obvia, entonces es:
¿cuál es la **mejor** forma de rellenar una tabla?

Es decir:
¿Qué hace que un agente sea bueno o malo, **inteligente** o no inteligente?

Métricas de desempeño

Es posible definir **varios agentes** para el mundo de la aspiradora simplemente rellorando la columna de la derecha de formas distintas.

Las **medidas de rendimiento** (**métricas de desempeño**) incluyen los criterios que determinan el éxito en el comportamiento del agente.

Cuando se sitúa un agente en un **medio**, éste genera una **secuencia de acciones** de acuerdo con las **percepciones** que recibe.

Esta secuencia de acciones hace que su hábitat pase por una **secuencia de estados**. Si la secuencia es la deseada, entonces el agente habrá actuado correctamente.

Métricas de desempeño

No hay una **única medida** adecuada para todos los agentes.

Se puede preguntar al agente por su **opinión subjetiva** acerca de su propia actuación, pero muchos agentes serían incapaces de contestar, y otros podrían engañarse a sí mismos.

Por tanto hay que insistir en la importancia de utilizar **medidas de rendimiento objetivas**, que normalmente determinará el diseñador encargado de la construcción del agente.

¿Cuál sería una **buena medida de desempeño** para la aspiradora autónoma?

Actividad con HackerRank

Actividad 5 - Bot saves princess



Actividad 5 - Bot saves princess

Entren a la siguiente liga:

<https://www.hackerrank.com/challenges/saveprincess>

Seguir las instrucciones y codificar en su lenguaje de programación favorito el **programa del agente** para que se mueva en su medio y cumpla su objetivo.

Racionalidad de un agente

La **racionalidad** en un momento determinado depende de cuatro factores:

- La **medida de rendimiento** que define el criterio de éxito.
- El **conocimiento del medio** en el que habita el agente.
- Las **acciones** que el agente puede llevar a cabo.
- La **secuencia de percepciones** del agente hasta este momento.

Racionalidad de un agente

Definición. Agente racional

En cada posible secuencia de percepciones, un agente racional deberá emprender aquella **acción** que supuestamente maximice su **medida de rendimiento**, basándose en las evidencias aportadas por la **secuencia de percepciones** y en el **conocimiento** que el agente mantiene almacenado.

Omnisciencia de un agente

Es necesario tener cuidado al distinguir entre **racionalidad** y **omnisciencia**.

Un **agente omnisciente** conoce el resultado de su acción y actúa de acuerdo con él; sin embargo, en realidad la omnisciencia no es posible.

Es decir, un agente omnisciente tiene racionalidad **perfecta**

Omnisciencia de un agente

La definición propuesta de racionalidad no requiere omnisciencia, ya que la elección racional depende sólo de la **secuencia** de percepción **hasta la fecha**.

Llevar a cabo acciones con la intención de **modificar percepciones futuras** se le denomina **recopilación de información**.

Esta es una parte importante de la racionalidad.

Otro ejemplo de **recopilación de información** lo proporciona la **exploración** que debe llevar a cabo el agente en un medio inicialmente desconocido.

Aprendizaje de un agente

La definición propuesta de racionalidad implica que el agente racional no sólo recopile información, sino que **aprenda** lo máximo posible de lo que está percibiendo.

La configuración inicial del agente puede reflejar un **conocimiento preliminar** del entorno.

A medida que el agente adquiere experiencia puede modificar y aumentar su conocimiento.

Autonomía de un agente

Se dice que un agente carece de **autonomía** cuando se apoya más en el conocimiento inicial que le proporciona su diseñador que en sus propias percepciones.

Un agente racional debe ser **autónomo**, debe saber aprender a determinar cómo tiene que compensar el conocimiento incompleto o parcial inicial.

Entorno de trabajo

Entorno de trabajo de un agente

Al definir un agente simple es necesario especificar las medidas de rendimiento, el entorno, los actuadores y sensores del agente.

A este conjunto de especificaciones se le llama **entorno de trabajo**.

Se utiliza el acrónimo **REAS** (Rendimiento, Entorno, Actuadores, Sensores) o **PEAS** (*Performance, Environment, Actuators, Sensors*).

*¿Cuál es el **entorno de trabajo** del agente en la Actividad 5?*

Entorno de trabajo de un agente

Tipo de agente	Medidas de rendimiento	Entorno	Actuadores	Sensores
Uber autónomo	Seguro, rápido, legal, viaje confortable, maximización del beneficio	Carreteras, otro tráfico, peatones, clientes	Dirección, acelerador, freno, señal, bocina, visualizador	Cámaras, sónar, velocímetro, GPS, tacómetro, visualizador de la aceleración, sensores del motor, teclado

Descripción REAS del entorno de trabajo de un Uber autónomo.

Propiedades de los entorno de trabajo

Totalmente observable vs parcialmente observable

Si los sensores del agente le proporcionan acceso al **estado completo del medio en cada momento**, entonces se dice que el entorno de trabajo es **totalmente observable**.

Un entorno puede ser **parcialmente observable** debido al ruido y a la existencia de sensores poco exactos o porque los sensores no reciben información de parte del sistema.

Propiedades de los entorno de trabajo

Totalmente observable vs parcialmente observable

¿El ejemplo de la Actividad 5 es **totalmente observable** o **parcialmente observable**?

Es totalmente observable.

¿El ejemplo del Uber autónomo es **totalmente observable** o **parcialmente observable**?

Es parcialmente observable.

Propiedades de los entorno de trabajo

Determinista vs estocástico

Un entorno es determinista o estocástico dependiendo del punto de vista del agente.

Si el próximo estado del medio está totalmente **determinado** por el estado actual y la acción ejecutada por el agente, entonces se dice que el entorno es **determinista**; de otra forma es **estocástico**.

Propiedades de los entorno de trabajo

Determinista vs estocástico

Un entorno presenta **incertidumbre** si **es parcialmente observable** o **no es determinista**.

Un agente no se tiene que preocupar de la **incertidumbre** en un medio **totalmente observable** y **determinista**.

Sin embargo, si el medio es **parcialmente observable** entonces puede parecer **estocástico**.

Propiedades de los entorno de trabajo

Determinista vs estocástico

¿El ejemplo de la Actividad 5 es **determinista** o **estocástico**?

Es determinista.

¿El ejemplo del Uber autónomo es **determinista** o **estocástico**?

Es estocástico.

Propiedades de los entorno de trabajo

Episódico vs secuencial

En un **entorno de trabajo episódico**, la experiencia del agente se divide en episodios atómicos.

Cada **episodio** consiste en la **percepción** del agente y la **realización** de una única acción posterior.

El **siguiente episodio** no depende de las acciones que se realizaron en **episodios previos**.

Propiedades de los entorno de trabajo

Episódico vs secuencial

En un **entorno de trabajo secuencial**, la decisión presente puede afectar a decisiones futuras.

Los medios episódicos son más simples que los secuenciales porque el agente no necesita tener un historial a través del tiempo.

Propiedades de los entorno de trabajo

Episódico vs secuencial

¿En el ejemplo de la Actividad 5 el entorno es **episódico** o **secuencial**?

Es episódico o secuencial dependiendo la implementación del programa del agente.

¿En el ejemplo del Uber autónomo el entorno es **episódico** o **secuencial**?

Es secuencial.

Un ejemplo de un entorno episódico es un clasificador de imágenes.

Propiedades de los entorno de trabajo

Estático vs. Dinámico

Si el entorno puede cambiar cuando el agente está deliberando, entonces se dice que el entorno es **dinámico** para el agente.

De lo contrario se dice que el entorno es **estático** para el agente.

Propiedades de los entorno de trabajo

Estático vs. Dinámico

Los medios estáticos son **fáciles** de tratar ya que el agente no necesita estar pendiente del mundo mientras está tomando una decisión sobre una acción, ni se preocupa por el paso del tiempo.

Si el entorno no cambia con el paso del tiempo, pero el rendimiento del agente cambia, entonces se dice que el medio es **semidinámico**.

Propiedades de los entorno de trabajo

Estático vs. Dinámico

¿En el ejemplo de la Actividad 5 el entorno es **estático** o **dinámico**?

Es estático.

¿En el ejemplo del Uber autónomo el entorno es **estático** o **dinámico**?

Es dinámico.

Propiedades de los entorno de trabajo

Discreto vs. Continuo

La distinción entre **discreto** y **continuo** se puede aplicar al **estado del medio**, a la forma en la que se maneja el tiempo y a las percepciones y acciones del agente.

Por ejemplo, un medio con estados discretos como el del juego del ajedrez tiene un número finito de estados distintos.

Propiedades de los entorno de trabajo

Discreto vs. Continuo

¿En el ejemplo de la Actividad 5 es **discreto** o **continuo**?

Es discreto.

¿En el ejemplo del Uber autónomo es **discreto** o **continuo**?

Es continuo.

Propiedades de los entorno de trabajo

Conocido vs. desconocido

Esta distinción no se refiere al ambiente en sí, sino al **estado de conocimiento del agente** (o diseñador) sobre las "leyes de la física" del entorno.

En un **entorno conocido** se dan los resultados (o probabilidades de resultado si el entorno es estocástico) para todas las acciones.

Propiedades de los entorno de trabajo

Conocido vs. desconocido

En un **entorno desconocido** el agente tendrá que aprender cómo funciona para tomar buenas decisiones.

La distinción entre **entornos conocidos** y **desconocidos** no es la misma que entre entornos **totalmente observables** y **parcialmente**.

Es muy posible que un entorno conocido sea parcialmente observable.

Propiedades de los entorno de trabajo

Conocido vs. desconocido

Por ejemplo, el juego de cartas de solitario, un agente conoce las reglas pero todavía no puedo ver las cartas que aún no han sido entregadas.

Por el contrario, un **entorno desconocido puede ser totalmente observable**: en un nuevo videojuego, la pantalla puede mostrar todo el estado del juego, pero todavía no sé qué hacen los botones hasta que los pruebo.

Propiedades de los entorno de trabajo

Conocido vs. desconocido

¿En el ejemplo de la Actividad 5 es **conocido** o **desconocido**?

Es conocido.

¿En el ejemplo del Uber autónomo es **conocido** o **desconocido**?

Es desconocido.

Propiedades de los entorno de trabajo

Agente individual vs multiagente

Dimos la definición de agente, pero no se ha explicado **qué entidades** se deben considerar agentes.

Ejemplo 1. Uber autónomo:

Sea el agente A un Uber autónomo.

Sea B otro vehículo dentro del ambiente de A.

¿Debe tratar el agente A al vehículo B como un agente, o puede tratarse como un objeto con un comportamiento estocástico?

Propiedades de los entorno de trabajo

Agente individual vs multiagente

La distinción clave está en identificar si el **comportamiento** de B está descrito por la maximización de una **medida de rendimiento** cuyo valor depende del **comportamiento** de A.

Ejemplo 2. Ajedrez:

Sea el agente A un jugador de ajedrez.

Sea B el jugador oponente.

El jugador oponente B intenta maximizar su medida de rendimiento, la cual, minimiza la medida de rendimiento del agente A.

Propiedades de los entorno de trabajo

Agente individual vs multiagente

El ajedrez es un entorno **multiagente competitivo**.

Regresando al Ejemplo 1.

Evitar colisiones maximiza la medida de rendimiento de todos los agentes, así que es un entorno **multiagente parcialmente cooperativo**

Además, compiten por maximizar su ganancia por lo que compiten por los clientes por lo que también es un entorno **parcialmente competitivo**.

Propiedades de los entorno de trabajo

Agente individual vs multiagente

¿En el ejemplo de la Actividad 5 es un entorno con **agente individual** o **multiagente**?

Es agente individual.

¿En el ejemplo del Uber autónomo es un entorno con **agente individual** o **multiagente**?

Es multiagente.

Recapitulando

	<u>Actividad 5</u>	<u>Uber autónomo</u>
Totalmente observable vs parcialmente observable	Totalmente observable	Parcialmente observable
Determinista vs estocástico	Determinista	Estocástico
Episódico vs secuencial	Episódico/Secuencial*	Secuencial
Estático vs dinámico	Estático	Dinámico
Discreto vs continuo	Discreto	Continuo
Conocido vs desconocido	Conocido	Desconocido
Agente individual vs multiagente	Agente individual	Multiagente

**Dependiendo la implementación del programa del agente*

Referencias bibliográficas

Schalkoff, R.J. (2011). Intelligent Systems: Principles, Paradigms, and Pragmatics. Sudbury, MA: Jones & Bartlett Learning.

Russell, S. J. and Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed). Pearson Education Limited.



ISyTE

Dr. Héctor Zatarain Aceves

Email: hector.zatarain@uabc.edu.mx

Teléfono: (646) 152 8244 Ext. 64350