

# CONSCIENCIA

La consciencia es un fenómeno tan confuso que se suele dar como inentendible desde el punto de vista científico y filosófico. Mas allá de las definiciones que se le pueda dar, parece haber consenso, no en qué significa sino en qué entidades son o no son conscientes. Pocas personas van a estar seguros de qué es la consciencia, pero van a estar seguros de que una computadora no es consciente y un humano si lo es. Es incluso una forma de pensar común el asumir que solo los seres humanos poseen consciencia, el fallo no es malentender el concepto de consciencia sino mas bien el asumir que es una propiedad intrínseca a entes específicos, sin justificación alguna.

Este fenómeno no es una propiedad que dota a la materia de cualidades específicas, hay que comprender que la materia interactúa de una manera determinada por su constitución física, y de ese comportamiento identificamos cualidades las cuales son abstractas y sujetas a nuestra forma de lenguaje y comprensión del mundo. En otras palabras, un ente no actúa a causa de que es consciente, sino que lo identificamos como consciente en base a cómo actúa. Esto aplica para cualquier adjetivo o clasificación que se le pueda atribuir a un fenómeno/ente.

Cuando comenzamos a pensar en qué es la consciencia, es probable que se piense primero en la percepción, en recibir información del entorno, tener sentidos con los cuales percibir el mundo que te rodea. Sin embargo parece obvio que una cámara de fotos o un micrófono no son entes conscientes a pesar de tener la capacidad de recibir información directa del entorno, al igual asumimos que un ojo humano no es consciente por si solo.

Pero no es lo mismo percibir que comprender el entorno, los seres humanos no solo recibimos la luz que rebota en nuestros ojos sino que identificamos cada cosa y la adjudicamos a conceptos abstractos e ideas, eso es la comprensión que se menciona. Sin embargo parece obvio que inteligencias artificiales capaces de reconocer los patrones en la información, identificar objetos y relacionarlos con conceptos del lenguaje natural no son conscientes según el consenso.

¿Entonces se trata de la subjetividad? ¿La consciencia se fundamenta en que haya una verdad única para el ente basada en su experiencia y sus aprendizajes?. La subjetividad es carencia de información, siempre que una inteligencia aprenda de cualquier conjunto de datos su entendimiento de lo que es verídico o lo que es falso va a estar totalmente influenciado por el enfoque de los datos de aprendizaje. Solo podemos saber lo que conocemos, no hay forma de que sea diferente. Cualquier inteligencia artificial contemporánea es totalmente subjetiva y no por ello asumimos que sea consciente.

Es algo común relacionar consciencia con creatividad e imaginación, esa capacidad que tenemos de crear cosas de la nada. Pero el punto es ese, no hay forma posible de crear algo de la nada, la imaginación y creatividad consiste en utilizar los patrones aprendidos de nuestra experiencia vital para combinarlos en una sola pieza, lo que llamamos arte es solo un collage de patrones previamente identificados en la información sensorial que recibimos a lo largo de nuestra vida, es parte de los procesos algorítmicos de inteligencia. A día de hoy existen inteligencias artificiales capaces de imaginar y crear arte previamente inexistente, no por ello las damos por conscientes.

¿Y la voluntad?, una IA contemporánea solo reacciona a entradas y está obligada a dar una salida, es un cálculo extenso, no tiene forma de actuar en base a sus deseos propios o en base a lo que considera necesario. Nace al darsele una entrada, y muere al devolver una salida. Para hablar de esto hay que explicar lo que es la voluntad en los seres que consideramos conscientes, nosotros los humanos actuamos en persecución de objetivos autoimpuestos fundamentados por sensaciones gratificantes y expectativas, corremos hacia lo que consideramos bueno y huimos de lo que consideramos malo, siendo el bien y el mal una serie de patrones que identificamos en vivencias pasadas. Si experimentamos un estímulo gratificante al vivir aquella situación en el pasado, vamos a relacionar cada propiedad identificada en ese escenario con algo bueno, y nuestro concepto de progreso y bienestar va a actualizarse en base a ello. El actuar de los humanos y animales se puede reducir a ejecutar acciones que inteligentemente relacionamos con escenarios que relacionamos con estímulos gratificantes. De este modo, cualquier agente que ejecute un aprendizaje por refuerzo realmente está actuando voluntariamente en persecución de aquello que considera necesario y de manera individual al igual que nosotros los seres "conscientes".

Entonces tal vez se trata de los sentimientos, ¿Son los sentimientos los que nos diferencian de el resto del universo?. Nuevamente pensemos en qué son los sentimientos. Los sentimientos/emociones son mecanismos físicos dentro de nuestro cerebro que sirven como referencias, como mencioné anteriormente, el actuar se sustenta en perseguir sentimientos buenos, si no hay sentimientos el cuerpo no tiene manera de identificar qué es lo que hay que hacer o no hay que hacer, el comportamiento sería errático y una especie animal de este tipo no podría sobrevivir de ninguna manera. Justamente aquellas sensaciones buenas son las relacionadas a la supervivencia, y las malas están relacionadas a escenarios con patrones relacionados a el peligro, lo bueno y malo está absolutamente influenciado por esas configuraciones implícitas en nuestra genética. Es un mecanismo replicable, simplemente un valor de referencia para determinar el actuar del agente. Tal como la voluntad, estos "sentimientos" están presentes en algoritmos de aprendizaje por refuerzo, donde el agente actúa basándose en la recompensa la cual representa sentimiento.

Ya solamente queda pensar en el libre albedrío. Una maquina siempre va a estar programada, determinada, son solamente componentes que interactúan de una forma ordenada y no hay forma de que tengan libre albedrío. Es común caer en este pensamiento, dando por hecho que los seres humanos tenemos libre albedrío y no estamos determinados por nuestra constitución genética o las leyes de nuestro universo. El comportamiento humano no es impredecible, es impredecible para nuestra tecnología actual. No conocemos el cerebro en su totalidad, pero eso es irrelevante para el tema de discusión porque es un hecho que almenos en esta escala de la realidad las interacciones son deterministas, las neuronas interactúan siguiendo unas leyes invariables, la materia actúa hoy igual que hace millones de años, incluso los sistemas mas absurdamente complejos no tienen margen de aleatoriedad en esta escala, una computadora con millones y millones de transistores que miden nanómetros ejecutando una simulación de millones de partículas puede ejecutar la misma simulación cada vez y presentar los mismos exactos resultados. Nuestro libre albedrío es una ilusión que actúa como mero espectador mientras las piezas del dominó caen por la inalterable ley de gravedad.

¿Entonces? ¿Qué es la consciencia?

Durante los cuestionamientos anteriores no se concluyó una definición de consciencia, esto es porque partimos de asumir que la consciencia existe y es necesaria de alguna forma, asumimos que nosotros somos conscientes y asumimos que hay cosas que no lo son, queremos ajustar la definición de tal forma que solo el humano sea considerado como ser consciente, sin embargo no hay cualidad presente en el ser humano que no pueda ser replicable de manera exacta o análoga.

Dejemos de pensar en ajustar la definición a lo que nos convenga. Existe un sistema que presenta todas las cualidades que se suelen identificar en un ser "consciente". Y podemos construirlo en base a los cuestionamientos dados anteriormente.

## ALGORITMO UPSILON

Voy a plantear una consciencia como un bucle continuo, un algoritmo que se repite a una frecuencia determinada. Los pilares fundamentales de esta consciencia son:

- La información sensorial (Estímulos)
- El flujo del tiempo (Continuidad del bucle)
- Las sensaciones (Calificaciones)
- La inteligencia
- Los actuadores

Los estímulos sensoriales van a ser procesados por una red neuronal, se van a reconocer patrones que relacionen cada propiedad de la información percibida, lo que el agente puede ver con lo que puede escuchar, relacionar la manera visual en la que su cuerpo se mueve al ejecutar movimientos voluntarios, identificar la individualidad física de los objetos en base a su manera independiente de movimiento y comportamiento, esto es el entendimiento del entorno desde cero.

Esta inteligencia a tiempo real es el verdadero reto de lograr construir una consciencia artificial, una red neuronal que opera a varios ciclos por segundo, una herramienta absolutamente poderosa para encontrar patrones en la información sensorial cruda. El enfoque actual de redes neuronales es totalmente digital y trae consigo un límite intrínseco a la computación contemporánea.

Este límite de la computación reside en que dentro de un software no existen las reacciones entre entidades. Todo actúa de manera separada, cada neurona o capa es algo por lo que hay que iterar activamente para que las operaciones den lugar. Dentro de los softwares programados no hay reacciones, un objeto virtual debe estar en constante estado de comprobación de su entorno para decidir activamente como comportarse, esto difiere de la realidad, donde cada objeto no actúa activamente sino que reacciona pasivamente a su entorno sin recibir información directa de este.

Hay que entender que las entidades separadas no existen como tal en nuestro mundo. Crear una virtualidad basada en partes que interactúan activamente entre si no es para nada análogo a la realidad, es mas bien análogo al lenguaje humano. Si se piensa a una partícula como entidad independiente dentro de un programa, vas a tener limitaciones en la cantidad de partículas que se pueden simular. Pero si se piensa a las partículas como parte de algo mas fundamental, como un fenómeno emergente de otro fenómeno mas básico, entonces la imposibilidad de un sistema de infinitas partículas se desvanece. Esto es solo un ejemplo de lo carente de sentido que suele ser crear sistemas donde se programan directamente comportamientos en lugar de lograr que emerjan indirectamente a partir de fenómenos fundamentales.

Las computadoras mas comunes en la actualidad son multi-propósito, tienen un lenguaje y toda la información contenida esta diseñada para ser totalmente variable. Cada byte de información es una representación binaria simbólica, las cantidades no existen físicamente en la electrónica de la computadora. Es por esto que hay una limitación significativa en la lógica de programación.

Una red neuronal capaz de operar de inicio a fin en milésimas de segundo o menos debe ser estrictamente un mecanismo físico que aproveche la reacción directa. Las neuronas no van a actuar cuando se las procese, las neuronas van a reaccionar instantáneamente a las entradas dadas. Tan pronto como se reciba una entrada se va a dar una salida, a la maxima velocidad permitida por el conductor eléctrico. La programación es la constitución física, no hay lenguaje ni abstracción.

Una vez superado el filtro de la inteligencia a tiempo real, hay que dar el siguiente paso, el cual es adjudicar a cada patrón reconocido una sensación a la cual voy a llamar calificación, porque justamente califica como bueno o malo a un patrón desencadenado por un conjunto de estímulos sensoriales crudos. Ciertas situaciones y escenarios presentan características, a esas características me refiero con patrones, patrones que se van a identificar como buenos o malos, esta calificación inicialmente va a ser instintiva, es decir, los lineamientos a tener en cuenta para determinar si un patrón es positivo o negativo están implícitos en el sistema de manera premeditada para incentivar al agente a actuar en persecución de un objetivo específico. Este objetivo es análogo al "objetivo" implícito de sobrevivir presente en nosotros los animales.

Sin embargo hay que tener en cuenta que lo que se considera como positivo o negativo va a cambiar a lo largo de la experiencia propia del agente. Este último tiene que almacenar en su memoria los patrones y su respectiva calificación, cada recuerdo relacionado a una sensación y una marca cronológica que le permita ubicarlo en el tiempo. Esto es lo que conforma la experiencia del agente y lo que va a ser útil para determinar qué patrones merecen mas peso positivo o negativo a la hora de actuar. Si lo pensamos bien, parece que depende en cierta medida de la repetición, es decir, si en algún momento de nuestra experiencia vital sentimos una sensación extremadamente gratificante, vamos a intentar replicar los patrones que identificamos en la situación que desencadenó dicha sensación potente. Inicialmente la gratificación va a ser mayor y mayor a medida que logremos repetir la sensación logrando replicar la situación, sin embargo si repetidamente logramos llegar a esa situación nos vamos a acostumbrar, y esos patrones van a reducir su peso en calificación.

Esta configuración tambien influye directamente en la forma de actuar del agente, es decir, se puede determinar en su constitución que la repetición de sensaciones gratificantes solo logre aumentar el peso de esa sensación y la relevancia de las situaciones relacionadas a ella, sin embargo esto probablemente estancaría al agente en una sola actividad la cual nunca le cansa.

Todo el actuar del agente consiste en intentar maximizar el valor de calificación de su situación actual, esto lo va a lograr mediante la inteligencia, la cual le va a permitir relacionar sus acciones con las reacciones en su entorno, va a comprender lo que puede hacer en base a lo que tiene para intentar llegar a un escenario en las que identifique propiedades buenas según su criterio subjetivo formado através de su experiencia y sus instintos.

Un agente que ejecute este algoritmo va a ser capaz de aprender a hablar, relacionando los patrones sonoros de las palabras con los patrones visuales de cada objeto o situación. O incluso relacionando las palabras con las sensaciones que presenta, esto sería un entendimiento de conceptos profundos como la felicidad y la tristeza, la curiosidad o el aburrimiento.

Este agente tambien va a tener gustos propios y actividades que desarrolle por pasión, aquello en lo que haya obtenido una sensación de gratificación al lograr un resultado es lo que va a promover la persecución de dichos resultados. Va a presentar una verdad real para si mismo, basada en su experiencia individual y basado en los aprendizajes y relaciones que haya identificado.

Y tambien va a presentar un autoconcepto, va a comprender que es su cuerpo y solo su cuerpo el que actúa siguiendo su voluntad, va a identificar cómo era antes, va a poder imaginar como será despues, su inteligencia le brinda esa capacidad de predecir o soñar basandose en lo que considera probable dada su experiencia.

Este fue el planteamiento filosófico y lógico del algoritmo Upsilon. A partir de esto es que va a desarrollarse proxicamente una muestra práctica.