

# Audiovisión e Inteligencia Artificial

Alejandro Brianza

Universidad Nacional de Lanús

Buenos Aires, Argentina

alejandrobrianza@gmail.com

## ABSTRACT

En un escenario actual en el que los avances de la Inteligencia Artificial y su posterior utilización al servicio de diversos desarrollos audiovisuales es cada vez mayor, resulta pertinente entender las posibilidades de estas herramientas para repensar la práctica profesional, los procesos de producción e incluso cuestiones relacionadas a la poética y estética que implica su implementación.

Este artículo define las nociones básicas acerca de las cuestiones tecnológicas contextuales, como también un recorrido por cuatro proyectos puntuales, que implican aplicaciones en los campos de la escritura de guión, montaje, foley y procesamiento de audio y video en tiempo real.

## Palabras clave

Audiovisión; Inteligencia Artificial; Robots; Automatización.

## INTRODUCCIÓN

“La cuestión no es si las máquinas inteligentes pueden sentir alguna clase de emoción, sino si las máquinas pueden ser inteligentes sin emociones”. Esto postulaba el adelantado Marvin Minsky en el año 1986 [6], desafiando un poco la idea de que las máquinas puedan superarnos en determinadas tareas careciendo de sentimientos o emociones. Sin embargo, últimamente, de cara al desarrollo que ha tenido la inteligencia artificial [IA], esta pregunta que durante un tiempo pareció contentarse con respaldos como el de Minsky, está comenzando a resonar cada vez más: ¿podrán los robots robarnos los empleos? La respuesta puede parecer desalentadora: progresivamente, sí podrán. Lo bueno, es que este avance de las máquinas sobre los empleos del hombre no es parejo para todas las áreas y disciplinas, ya que hay tareas que son más fáciles de automatizar que otras. Este dato no es menor, ya que nos debería permitir estudiar cuáles serán los empleos que estarán disponibles en el futuro, para así formarnos y estar preparados para ese momento. □Estos planteos, que vienen dando vuelta hace años pero que cada vez se tornan más frecuentes, hacen tambalear desde la industria hasta la educación: podemos pensar que un servicio de vigilancia es más automatizable que el dictado de una clase. O la conducción de un camión de transporte, más que la construcción de una casa. Para ir al punto que nos interesa, repasaremos qué está pasando con los desarrollos de IA al

servicio del desarrollo de productos audiovisuales.

Pero vamos desde el principio, sabemos que nos rodea y que cada vez está más presente en nuestras vidas, pero ¿a qué nos referimos cuando hablamos de IA?

## QUÉ ES LA INTELIGENCIA ARTIFICIAL

Se puede decir que la IA es aquel mecanismo utilizado por máquinas –computadoras– que les permite discernir opciones frente a un escenario problemático y resolverlo de la mejor forma posible, en la mayoría de los casos de forma exitosa.

Si bien se considera a Alan Turing como el mayor impulsor de la idea de que las máquinas pueden ser inteligentes –e incluso diseñó en 1950 una prueba que actualmente se conoce como Test de Turing para ponerlas a prueba–, el concepto de inteligencia artificial como tal, surgió en 1956 en una reunión académica de verano impulsada por John McCarthy a la que también asistieron Marvin Minsky, Allen Newell y Herbert Simon. El encuentro tenía como objetivo estudiar las posibilidades de las máquinas acerca de la utilización del lenguaje verbal y establecer una serie de caminos viables por los cuáles se esperaba que avanzaría la tecnología. Aunque la serie de premisas que postularon en general no se cumplió, es innegable que esta reunión abrió un campo de desarrollo hasta ese entonces poco estudiado.

A partir de este hito, algunos programadores continuaron con lo planteado y trabajaron en convertir en código los procesos de comprensión de lenguaje; otros profundizaron en conocer los mecanismos de aprendizaje tanto en el ser humano como en distintas especies animales y, con el correr de los años, las tareas plausibles de ser realizadas por máquinas fueron siendo cada vez más.

Existen ciertas condiciones que favorecen el funcionamiento general de la IA, como las grandes cantidades de información a procesar para poder buscar patrones de comportamiento con mejores resultados, o la descentralización de los subsistemas que componen a la IA en cuestión, puesto que la suma de muchas tomas de decisiones a problemas simples suele tener mejor rendimiento que una sola toma de decisión para un problema complejo. Es por eso que actualmente, los desarrollos en el campo de la IA se aprovechan del big data y de los incontables terabytes de información que son generados por segundo por nosotros mismos en redes

sociales, computadoras personales o teléfonos celulares, para sacar conclusiones y operar, en el sentido que hayan sido programados. Las llamadas redes neuronales artificiales, filtran esta inmensa cantidad de datos y a medida que surgen conclusiones al respecto, los nodos y conexiones de la red se reacomodan, mejorando el rendimiento de interpretación para el futuro.

Ahora bien, por más que cumplan o se aprovechen estas condiciones favorables, cuando hablamos de IA, tenemos que tener presente que existen distintos niveles de procesamiento de información que pueden dividirse en las siguientes cuatro categorías:

### **1. Máquinas reactivas**

Esta es la primera categoría respecto a la capacidad de resolución de problemas. Se las llama máquinas reactivas precisamente porque su accionar depende únicamente de una reacción pautada a la recepción de ciertos estímulos para los que están programadas. No guardan recuerdos ni aprenden de las experiencias transitadas para sus decisiones futuras. □ Un histórico caso al respecto es el de la conocida Deep Blue, una computadora diseñada por IBM para jugar al ajedrez, que en el año 1997 logró derrotar al entonces campeón Garry Kasparov. La anécdota de aquel día, la relata él mismo a modo de crónica personal [4] con detalles acerca de cómo se sintió al competir con Deep Blue a los ojos de una inusual cantidad de público para lo que se acostumbra en una partida de ajedrez, conociendo la implacabilidad de la computadora y lo que significaba ese momento para el mundo entero.

La máquina conoce las piezas y sus posibilidades y funciona a partir del estímulo del adversario, calculando una cantidad dada de movimientos próximos posibles y eligiendo entre las posibilidades que tiene precargadas, la mejor movida.

Son máquinas que no sirven para otras tareas que para aquello que fueron diseñadas, y eso es lo que las vuelve estables y confiables en los campos que dominan, eligiendo la mejor opción entre un alto número de posibilidades, con alta probabilidad de éxito.

### **2. Máquinas con memoria limitada**

Estas máquinas, además de tener programado un abanico de respuestas posibles a los problemas que afronten, están preparadas para tomar múltiples datos del ambiente y, como su nombre ya lo anuncia, retenerlos por un corto lapso de tiempo para tomar decisiones en torno a ellos. A esta incorporación de información, se la combina con machine learning, un método estadístico que implica específicamente la capacidad de aprender o bien, adaptarse progresivamente en la realización de una tarea determinada sin que la máquina haya sido diseñada para ese fin. □ Esta es la tecnología que Google utiliza, por ejemplo, en Waymo [12] –su proyecto de automóviles autónomos– donde la computadora que hace funcionar el sistema tiene

precargados ciertos mapas y datos acerca de semáforos y señalizaciones, pero incorpora en tiempo real el tránsito, los comportamientos y tipo de manejo de los autos que tiene en proximidad. Al finalizar el viaje, estos últimos datos no se almacenan, ni generan un cambio de ningún tipo en la programación del auto. □ También se han experimentado máquinas con memoria limitada en muchos videojuegos que están pensados más que nada para ser jugados en modo multijugador, pero que en los modos para un solo jugador permiten la inclusión de bots –otros participantes virtuales que maneja la propia computadora– que funcionan con estas tecnologías: aprenden del entorno o escenario mientras dure la partida y al finalizar o reiniciar el juego vuelven a su estado inicial.

### **3. Máquinas con una Teoría de la mente**

Aquí cruzamos una línea entre lo hecho y lo no hecho, puesto que las máquinas que dominan la Teoría de la mente son, hoy por hoy, un deseo de quienes impulsan sus desarrollos. La denominada Teoría de la mente es una habilidad cognitiva que permite entender y predecir los comportamientos de un individuo gracias a la comprensión de sus estados de ánimo, ideas, pensamientos y emociones.

Así, las computadoras no solo tendrían una programación dada en adición a la memoria limitada, sino que estarían dispuestas a resolver los problemas poniendo como variable el estado de ánimo o las reacciones de su interlocutor.

### **4. Máquinas con conciencia propia**

El punto cúlmine, es conseguir que una computadora no solamente administre las emociones externas, sino que tenga las propias. Para los humanos, entender sensaciones como dolor o alegría, en parte es posible por haberlas experimentado y aprendido de ellas en el pasado. Así, los desarrollos que persiguen la generación de una conciencia artificial están relacionados con modelos de aprendizaje, pensamiento y memoria.

Aquí toma especial sentido la cita de Marvin Minsky [6] que da comienzo a este artículo: ¿pueden las máquinas ser inteligentes cuando son incapaces de sentir emociones? Hay planteos que pueden funcionar como respuesta a esta pregunta, tales como el que postula Yuval Harari [2], quien piensa a la inteligencia humana como una suma de algoritmos biológicos inscriptos dentro de la evolución de la especie. Esto implica que de poder descifrarse esos algoritmos –un proyecto tan ambicioso como ha sido en años anteriores el de resolver la configuración del genoma humano–, podrían tranquilamente recrearse artificialmente dentro de una máquina. □

Al igual que las máquinas que dominan la Teoría de la mente, por ahora solo hemos conocido a estas computadoras en películas de ciencia ficción.

## INTELIGENCIA ARTIFICIAL APLICADA

Si bien dos de las cuatro categorías están aún en desarrollo, las máquinas reactivas y las que cuentan con memoria limitada avanzan a pasos agigantados. Ahora que conocemos a qué nos enfrentamos, podemos avanzar en el planteo inicial y abordar el campo audiovisual. Les comparto cuatro ejemplos de aplicaciones bastante novedosas, que nos van a permitir tener una idea de cuáles son las posibilidades que existen hoy en día a partir de intentar cubrir distintas necesidades de las etapas de producción audiovisual.

### Sunspring

Oscar Sharp y Ross Goodwin quisieron desafiar los límites de la IA y construyeron un software que fuera capaz de redactar guiones a partir de otros preexistentes. Así, le dieron vida a Benjamin, una IA funcionando sobre una red neuronal LSTM –de las siglas en inglés long short term memory–, a la que cargándole una gran lista de guiones de películas, es capaz de generar uno nuevo. La base de datos constó de títulos tan diversos como Alien, Blade Runner, Cazafantasmas, Armageddon, G.I. Joe, Jurassic Park, Hombres de Negro y Star Trek, entre otros cientos.

Con Benjamin bajo el brazo, se anotaron en Sci Fi London, un concurso que dispone de una categoría en la que solo se cuenta con 48 horas para grabar y postproducir un cortometraje. Rodaron Sunspring [10], obviamente, poniendo a prueba un guión que Benjamin les ofreció, quedando increíblemente en el top 10 del concurso. □Una vez terminado el certamen, los autores dieron a conocer su herramienta convirtiéndose en blanco de las más altas críticas en cuestiones éticas para con el concurso; pero también de diversidad de halagos hacia la capacidad de Benjamin para resolver el guión del corto. De hecho, Sunspring se volvió famosa y viral en redes, incluso por encima de los audiovisuales que ganaron la categoría. □¿Es justo que hayan utilizado IA? Sharp y Goodwin alegan que las reglas del concurso no lo prohibían.

### Morgan

Morgan es la típica película de ciencia ficción sobre IA: sus posibilidades, la relación con los humanos y tantas otras cosas que vimos ya en distintas historias. Sin embargo, una anécdota particular la convirtió en especial: desde FOX –productora de la película– desafiaron a IBM para que Watson [3], un desarrollo de IA de su propiedad, proponga el montaje para el tráiler. Si bien Watson nunca había trabajado en postproducción audiovisual, en IBM aceptaron el reto y comenzaron a convertir el largometraje en datos analizables. El poder reconocer puntos de tensión, protagonistas, personajes de relleno y diálogos narrativamente importantes, entre otras variables, permitió que Watson elija cuidadosamente una selección de escenas que deberían mostrarse para la promoción de la película. □No solo los resultados son dignos de competir en calidad

con cualquier otro tráiler [7], sino que de no saber el detrás de escenas de su postproducción, nunca nos daríamos cuenta, como espectadores, que ese tráiler fue montado por una IA.

### Visually Indicated Sound

La técnica de foley es aquella que nos permite recrear en una situación de estudio aquellos sonidos y efectos que por una cuestión u otra, no pudimos o no estaba planeado registrar durante el rodaje. Así, se recrean desde pasos hasta complejas manipulaciones de objetos, intentando dar la máxima credibilidad para que estos sonidos doblados pasen desapercibidos al momento de ser montados con una secuencia visual. □En la industria, existen grandes estudios de foley que generan los sonidos necesarios, pero que también los almacenan en distintas librerías para reutilizarlos en ocasiones futuras. Podemos decir que la tarea de los profesionales de esta técnica se divide en dos: aquellos artistas de foley que recrean las situaciones deseadas en el estudio de grabación; y aquellos técnicos que conocen y dominan a la perfección la colección pregrabada que disponen, sabiendo qué sonido se puede adecuar a las situaciones a sonorizar artificialmente. □Esta segunda tarea, ocupa en general mucho tiempo de reescucha y selección hasta que se consigue el efecto o sonido deseado. Con esta problemática como norte, el profesor Andrew Owens y su equipo, desarrollaron en el MIT una aplicación que puede, a partir del ingreso de un video sin banda sonora, predecir qué sonidos foley de una gran librería que tiene guardada, coinciden mejor para sonorizar esa situación [8]. El software está en etapa experimental pero ha obtenido muy buenos resultados [11]. Tal es así que se ha probado en tests, comparando un video con su banda sonora original contra un video resonorizado por la IA de Owens, y al público le resulta en la mayoría de los casos igual de creíble uno que el otro.

### Leviatán

Leviatán es un proyecto de Andamio [1] pensado como performance en situación de concierto audiovisual, en el que interactúan un instrumentista tocando una flauta dulce Paetzold con música electroacústica y procesos audiovisuales en tiempo real. □Por un lado, se plantea un sistema musical interactivo que funciona sobre un patch de SuperCollider utilizando la técnica de machine listening –el machine learning aplicado a procesos sonoros–. Esto permite que la pieza se genere a sí misma tomando como materia prima distintos eventos generados por la flauta Paetzold, que ingresan al sistema mediante vía micrófono. □Los materiales visuales están pregrabados y son puestos en juego con CineVivo, siendo también adaptados y modelados por estos los eventos sonoros, construyendo una sinuosa narrativa dependiente de la interrelación entre performer, compositor y computadora.

La pieza plantea una duración flexible, estructurada en doce secciones que pueden durar un mínimo de treinta segundos y un máximo de infinito, teniendo la IA un comportamiento diferente cada parte. Es el performer quien dispone los cambios entre secciones: con un controlador que se comunica por protocolo OSC –un formato de mensajes informáticos, que cada vez se utiliza más como alternativa a la normativa MIDI, ya que dispone de más resolución y además puede ser transferido en tiempo real de manera inalámbrica en una red local–, envía las señales necesarias hacia SuperCollider y CineVivo, modificando así la forma de procesar sonido y visuales.



**Figura 1. Performance de Leviatán, de Andamio. Para flauta Paetzold y procesos audiovisuales en tiempo real.**

## COMENTARIOS FINALES

Con este panorama, ahora sí podemos regresar al inicio y preguntarnos: ¿podrán los robots robarnos los empleos? se me ocurren al menos dos caminos posibles para proponer respuestas. El primero es sobre un escenario más apocalíptico, en el que efectivamente, la IA se convierte en competencia para nosotros como profesionales.

La segunda opción, un escenario más integrado, nos permite pensar a estos desarrollos como una herramienta más para destinar a la producción en un sentido amplio: para Sharp y Goodwin fue una herramienta de preproducción que los ayudó a tener un guión; para IBM una de postproducción que los asistió con el montaje. Todo parece indicar que los empleos están en la zona segura siempre que impliquen el uso de la creatividad. Al menos por ahora, mientras las máquinas con conciencia propia no proliferen. □ La actualización profesional siempre fue una característica de esta y tantas otras profesiones, y entendiendo a los desarrollos de IA como herramientas, pueden tenerse en cuenta como algo más que aprender a utilizar. El resto es hacer futurología. ¿Será que se vienen los sindicatos anti IA?

## REFERENCIAS

1. Andamio. Web oficial. Recuperado el 20 de marzo de 2019 de <https://www.andamio.in/>
2. Yuval Noah Harari. 2015. *Homo Deus: A brief history of tomorrow*. Harper Collins Publishers.
3. IBM - Watson. Web oficial. Recuperado el 20 de marzo de 2019 de <https://www.ibm.com/watson?lnk=404>
4. Gary Kasparov. 2017. *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. PublicAffairs.
5. Ray Kurzweil. 2012. *How to Create a Mind*. Viking Penguin Press.
6. Marvin Minsky. 1986. *The society of mind*. Simon and Schuster.
7. Morgan | IBM Creates First Movie Trailer by AI [HD] | 20th Century FOX. Video. (31 de agosto de 2016). Recuperado el 20 de marzo de 2019 de <https://www.youtube.com/watch?v=gJEzuYynaiw>
8. Andrew Owens, Phillip Isola y colaboradores. 2015. *Visually Indicated sounds*. Recuperado el 20 de marzo de 2019 de <https://arxiv.org/pdf/1512.08512.pdf>
9. Eric Sadin. 2017. *La humanidad aumentada*. Caja Negra Editora.
10. Sunspring | A Sci-Fi Short Film Starring Thomas Middleditch. Video. (9 de junio de 2016). Recuperado el 20 de marzo de 2019 de <https://www.youtube.com/watch?v=0FW99AQmMc8>
11. Visually-Indicated Sounds. Video. (12 de junio de 2016). Recuperado el 20 de marzo de 2019 de <https://www.youtube.com/watch?v=0FW99AQmMc8>
12. Waymo. Tecnología. Recuperado el 20 de marzo de 2019 de <https://waymo.com/tech/>