COMILLAS
UNIVERSIDAD PONTIFICIA

Facultad de Derecho

DEEP LEARNING Y SUS APLICACIONES EN EL CAMPO DE LA CREACIÓN ARTÍSTICA

Autor: Isabel Pastor García
Director: Carlos Martínez de Ibarreta Zorita

MADRID | Abril 2022

Tabla de Contenidos

I. Introducción *5*

II. Marco teórico del Deep Learning *8*

1. Inteligencia artificial, *Machine Learning* y *Deep Learning* 8

2. Redes Neuronales Artificiales 12

2.1. Concepto 13

2.2. El perceptrón 13

2.3. El perceptrón multicapa – Redes neuronales multicapa 16

2.4. El proceso de aprendizaje en las redes neuronales artificiales 17

3. Redes Neuronales Profundas 19

3.1 Redes neuronales convolucionales 20

3.2 Redes neuronales recurrentes 26

III. Redes neuronales profundas en el campo de la creación artística *34*

1. *Deep Learning* en el análisis de obras de arte 34

2. *Deep Learning* en la edición de imágenes: 35

2.1 Google Deep Dream 35

2.2 Neural Style Transfer 36

3. Uso de redes neuronales recurrentes en la creación de textos 40

IV. Implicaciones de negocio de las técnicas de Deep Learning aplicadas al arte *45*

1. *Deep Learning* en la industria de las artes plásticas 45

2. *Deep Learning* en la industria de la creación audiovisual 46

3. *Deep Learning* en la industria musical 49

V. Reflexiones sobre el Deep Learning aplicado al arte *52*

1. ¿Son los modelos de *Deep Learning* capaces de generar arte? 52

2. Autoría y problemas éticos en la aplicación del *Deep Learning* al arte 54

2.1. Autoría y copyright de las obras generadas por inteligencia artificial 54

2.2. Datasets y sesgos implícitos 55

2.3 Aspectos negativos para los artistas: La lucha por la atención del espectador 57

VI. Conclusión *59*

Resumen

El *Deep Learning* se ha convertido en uno de los campos del *Machine Learning* que más se ha desarrollado en la última década, ofreciendo aplicaciones potencialmente revolucionarias para la mayoría de las industrias. Las industrias relacionadas con el arte no son una excepción. En este trabajo de investigación se estudia el *Deep Learning* desde un punto de vista técnico, tratando los conceptos teóricos más relevantes con el objetivo de comprender el funcionamiento del mismo. Además, se lleva a cabo una revisión del estado del arte en su aplicación a la creación artística. Sin embargo, la implementación de este tipo de tecnologías no sólo afecta al artista en tanto a sujeto creador de forma individual, sino que existen grandes implicaciones de negocio de la aplicación de este tipo de técnicas, tanto en la industria de las artes plásticas, como en la industria cinematográfica y en la industria musical. Además, la llegada del *Deep* Learning al arte no está exenta de dificultades en el plano ético y jurídico, problemáticas a las que no hay una respuesta unánime en el plano internacional. En definitiva, la aplicación del *Deep Learning* al campo artístico abre una infinidad de oportunidades en la industria, contando con un potencial transformador innegable que afectará a la mayoría de los actores de la misma.

Palabras clave:
Deep Learning, redes neuronales artificiales, creación artística, industria de las artes plásticas, industria cinematográfica, industria musical.

Abstract

Deep Learning has become one of the most rapidly developing fields of Machine Learning in the last decade, offering potentially revolutionary applications for most industries. Art-related industries are no exception. In this paper we will explore *Deep Learning* from a technical and computational point of view, analyzing the key concepts that make up the theoretical backbone of deep neural networks. In addition, a review of the state of the art in its application to artistic creation is carried out. However, the implementation of this type of technology not only affects the artist as an individual creator, but there are great business implications of the application of such techniques, both in the visual arts, film, and music industries. In this regard, we will analyze some of

the most relevant business opportunities derived from *Deep Learning* in the aforementioned industries. In addition, the deployment of Deep Learning techniques in art is not free from ethical and legal difficulties, issues to which there is no unanimous response at the international level and which will be explored in this paper. In short, the application of Deep Learning to the artistic field opens up an infinite number of opportunities in the industry, with an undeniable transformative potential that will affect most of its players.

Keywords:
Deep Learning, artificial neural networks, artistic creation, art industry, film industry, music industry.

I. Introducción

Imagina caminar los pasillos de uno de los museos más prestigiosos del mundo. Cuadros de Tiziano, Da Vinci y Van Gogh llenan las paredes. De repente, el retrato de un hombre con ropajes antiguos capta tu atención. Los intensos tonos azules te impresionan, por lo que decides comprobar el nombre del artista. Sin embargo, descubres que no ha sido elaborado por un artista humano, sino por un modelo que aplica técnicas de *Deep Learning*. Esto no es un supuesto teórico, o una previsión futurista. En junio de 2019, la compañía francesa Obvious llevó a cabo una exposición en el museo Hermitage de San Petersburgo. Todas las obras expuestas fueron generadas por redes neuronales profundas. Es decir, no existe artista ni pincel, sino que la actividad creativa es llevaba a cabo por un algoritmo.

Pudiera parecer que el mundo del arte –ya sean artes plásticas, el mundo de la música o incluso la palabra escrita– está muy alejado de la tecnología. Conceptualmente parecen estar en las antípodas, pero a través de la presente investigación se comprobará cómo las técnicas de *Deep Learning* consiguen cerrar la brecha existente, así como estudiar de qué forma este tipo de técnicas pueden revolucionar el arte en tanto a nivel disciplina como a nivel industria. El objetivo de este trabajo es hacer una revisión del estado de arte de las redes neuronales profundas o *Deep Learning* desde un punto de vista técnico, así como su aplicación en el campo de la creación artística. También se analizará las implicaciones de negocio de la aplicación del *Deep Learning* y las cuestiones éticas que ésta suscita.

En la última década –y especialmente en los últimos cinco años– el *Deep Learning* ha sido objeto de profusas investigaciones desde el mundo académico. Las infinitas posibilidades que parecen ofrecer los modelos que usan este tipo de técnicas hacen que, sin duda, estemos ante uno de los temas más candentes en el campo de la inteligencia artificial. El *Deep Learning* es aplicable a la gran mayoría de campos, siendo el artístico el escogido para esta investigación. La selección del campo artístico se basa en el hecho de que a priori el arte se relaciona en el imaginario colectivo con una creatividad orgánica, monopolio de los artistas humanos que expresan sus sentimientos a través de su obra. Por ello, he considerado especialmente interesante unirlo con el mundo tecnológico –en concreto, modelos capaces de crear piezas artísticas.

Así las cosas, considero que poner en relación uno de los temas que más expectación causa entre los académicos que estudian *Machine Learning* con las artes constituye una temática interesante a tratar en un Trabajo de Fin de Grado. Además, se ha seleccionado como tema desde el firme convencimiento de que el *Deep Learning* es y será toda una revolución para un gran abanico de industrias, siendo las industrias de las artes plásticas, la musical y la cinematográfica algunas de las que serán objeto de estudio en este trabajo.

En cuanto a la metodología seguida, se ha llevado a cabo una revisión del estado del arte de las técnicas de *Deep Learning* a través de la consulta de artículos académicos, libros y publicaciones online. Cabe destacar que el segundo capítulo del trabajo ha sido elaborado junto a mi compañera Laura García Tudela, pues se nos recomendó por parte de nuestro tutor realizar una investigación conjunta sobre el marco teórico del *Deep Learning.* Sin duda, llevar a cabo una investigación conjunta ha sido una decisión fructífera, pues ha facilitado la comprensión de los conceptos y la profundización en el plano teórico de la materia.

El trabajo consta de seis capítulos, siendo el primero el presente, dedicado a realizar una introducción a la investigación realizada.

El segundo capítulo del trabajo desarrolla el marco teórico del *Deep Learning*. En él, se diferencian los conceptos de *Deep Learning, Machine Learning* e *Inteligencia Artificial*. Después, se abordarán los distintos tipos de aprendizaje y la arquitectura básica de las redes neuronales artificiales, desde el perceptrón hasta las redes neuronales multicapa. A continuación se desarrollará el proceso de aprendizaje en las redes neuronales profundas, destacando los conceptos de computación *feed forward* y *backpropagation*. Por último, se analizarán las redes neuronales profundas, en especial las redes neuronales convolucionales y las redes neuronales recurrentes. En las redes neuronales convolucionales se incidirá en el funcionamiento de cada una de sus capas. En el caso de las redes neuronales recurrentes, se desarrollará cómo se consigue añadir el componente secuencial que caracteriza a este tipo de redes, notablemente a través del *backpropagation through time* y las *gated cells.*

En el tercer capítulo se abordará la aplicación del *Deep Learning* en el campo de la creación artística. En primer lugar, se expondrá de qué manera se pueden usar las técnicas

comentadas para analizar obras de arte. Después, se estudiarán las aplicaciones del *Deep Learning* en la edición de imágenes, en concreto *Google Deep Dream* y *Neural Style Transfer* –aplicado tanto a imágenes como a su puesta en común con piezas musicales. Finalmente se desarrollará el *Deep Learning* en la creación de textos, exponiendo ejemplos como la creación de poemas, la descripción de imágenes y la generación de guiones de cine.

En el cuarto capítulo se explorarán las implicaciones de negocio de las técnicas de *Deep Learning* aplicadas al arte. En concreto, se estudiarán tres de las grandes industrias del campo artístico: las artes plásticas, la creación audiovisual y cinematográfica y, finalmente, la industria musical. En el marco de la industria de las artes plásticas se mostrarán empresas dedicadas a la generación de cuadros, así como aplicaciones que ayudan a los inversores en sus decisiones de compra de obras de arte. Respecto a la industria de la creación audiovisual, se explorarán las oportunidades de negocio que implica el *Deep Learning* tanto en la fase de preproducción como postproducción, siendo la generación de trailers, el análisis de guiones y el desarrollo de efectos especiales algunas de las cuestiones a tratar. Finalmente, en el campo de la industria musical se estudiarán empresas que centran su actividad en la generación de piezas musicales, la detección de música y la masterización de audio.

Respecto al quinto capítulo, se expondrán una serie de reflexiones acerca de la aplicación del *Deep Learning* al campo del arte. En particular, se abordará la cuestión de la validez de las piezas generadas por modelos de *Deep Learning* en tanto que obras de arte. Después, se tratarán los problemas en términos de derechos de autor que la generación de obras puede acarrear, estudiando la respuesta que se le ha dado a la citada problemática en distintos ordenamientos jurídicos. A continuación, se analizarán ciertas cuestiones de carácter ético y legal sobre la selección de los datasets usados para entrenar los modelos de *Deep Learning*. Finalmente, se tratarán los posibles efectos negativos de la aplicación de técnicas de *Deep Learning* para los artistas, especialmente en términos de empleo y de lucha por la atención del espectador.

El sexto capítulo consistirá en la exposición de las conclusiones dilucidadas a lo largo del trabajo de investigación.

II. Marco teórico del *Deep Learning*

En este primer capítulo vamos desarrollaremos el marco teórico-conceptual del *Deep Learning*, con el objetivo de comprender cómo funcionan desde las redes neuronales más simples hasta los distintos tipos de redes neuronales profundas.

Comenzaremos por trazar una diferenciación conceptual entre inteligencia artificial, *Machine Learning* y *Deep Learning*. Después, nos adentraremos en el campo de las redes neuronales artificiales, analizando sus formas más simples –el perceptrón–, para luego estudiar formas más complejas como las redes neuronales multicapa. En este sentido, analizaremos el proceso de aprendizaje de las redes neuronales artificiales, explorando conceptos tan importantes como la computación *feed forward* y la retropropagación del error o *backpropagation*. Finalmente, analizaremos las redes neuronales profundas, centrándonos en dos de los tipos más comunes: las redes neuronales convolucionales –especialmente apropiadas para el tratamiento de imágenes– y las redes neuronales recurrentes –muy útiles para tareas con un componente secuencial o temporal. En el caso de las redes neuronales convolucionales, estudiaremos cómo funcionan cada una de sus capas –capas convolucionales, de *pooling* y *fully connected layers*– para llevar a cabo una potentísima labor de reconocimiento de imágenes. Respecto a las redes neuronales recurrentes, analizaremos cómo añadir el componente secuencial a las redes neuronales, desarrollando conceptos como el *backpropagation through time* y las *gated cells*.

1. Inteligencia artificial, *Machine Learning* y *Deep Learning*

Las técnicas de *Deep Learning* han supuesto una verdadera revolución en el campo de la inteligencia artificial. En la última década, y especialmente en el último lustro, dichas técnicas han conseguido resolver problemas que antes resultaban inalcanzables para las máquinas, siendo grandes ejemplos el reconocimiento de imágenes al nivel de los humanos, asistentes digitales (Amazon Alexa) y los coches de conducción autónoma. El desarrollo del *Deep Learning* ha afectado a la práctica totalidad de las industrias en una gran cantidad de aspectos, y a día de hoy parece ser más relevante que nunca.

Sin embargo, antes de adentrarnos en el mundo del *Deep Learning*, es necesario explorar una serie de conceptos previos, notablemente Inteligencia Artificial, aprendizaje supervisado, no supervisado y mixto, y *Machine Learning*.

Existe una estrecha relación entre los conceptos de Inteligencia Artificial, *Machine Learning* y *Deep Learning.* En concreto, se trata de subgrupos:

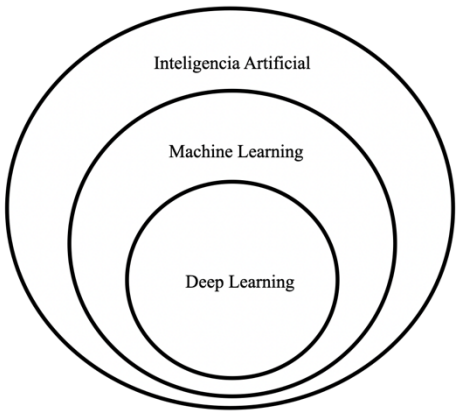


Figura 1
Esquema IA, ML y DL
Elaboración propia

Como se puede observar en la figura 1, la Inteligencia Artificial contiene los conceptos de *Machine Learning* y *Deep Learning*, pero abarca mucho más que eso. El concepto de Inteligencia Artificial data de la década de los cincuenta, con la idea de automatizar tareas que normalmente son realizadas por humanos (Chollet, 2018). En concreto, el término se acuñó en *Dartmouth College* , y se desarrolló con la voluntad de identificar los procesos formales que integran los comportamientos humanos a nivel de inteligencia en una gran variedad de campos, desde el ajedrez hasta el procesamiento de lenguaje.

El concepto de *Machine Learning* vivió su auge a partir de la década de los 90, siendo más específico que el concepto de Inteligencia Artificial, pues se refiere a aquellos modelos que no son programados con reglas diseñadas por humanos, sino que se entrenan. Dichos modelos llevan a cabo labores predictivas –clasificación o regresión–, utilizando para ello un gran abanico de algoritmos, siendo algunos de los más comunes SVM[1], árboles de decisión, PCA[2], *K-means Clustering* y las redes neuronales. Dentro del

[1] Support Vector Machine
[2] Principal Component Analysis

Machine Learning encontramos distintos métodos de aprendizaje, los cuales se agrupan normalmente en dos categorías: supervisado y no supervisado.

En primer lugar analizaremos los modelos de aprendizaje supervisado. El propósito es dar al modelo los datos (*inputs*), las respuestas (*outputs*) para que el modelo aprenda los patrones y las reglas que los relacionan de forma autónoma, para luego poder aplicarlo a otros *datasets* de los cuales desconocemos las respuestas de cara a poder realizar predicciones. Dicho de otra forma, el modelo aprende a través de ejemplos. En este respecto, es fundamental aplicar medidas de performance para determinar el nivel de error en el modelo, y, por tanto, la calidad predictiva del mismo.

De esta forma, tres son los elementos básicos de los modelos de *Machine Learning* que aplican aprendizaje supervisado (Tamir, 2020):

1. Proceso de decisión: A través de diversos algoritmos que normalmente realizan labores de clasificación o predicción, el modelo trata de hallar los patrones o las reglas internas presentes en los datos que son clave para realizar la tarea.

2. Función de error: Con la intención de conocer la calidad predictiva o clasificatoria del modelo. Algunos ejemplos de funciones de pérdida son RMSE[3], MAPE[4], MAE[5] (para modelos de regresión) y entropía cruzada binaria o multiclase y *hinge loss* (para modelos de clasificación).

3. Proceso de optimización del modelo: A la vista del error detectado, se ajusta y actualiza el modelo para mejorarlo.

En segundo lugar, los modelos de aprendizaje no supervisado se centran en los *inputs*. Es decir, no se sirve de *outputs* o variables respuesta sino de los datos en sí en aras de comprenderlos y entender las correlaciones, estructuras y patrones existentes en los mismos. Frecuentemente se usa con propósitos relacionados con la visualización de datos y eliminación de ruido y compresión en los mismos. En muchas ocasiones se usa como

[3] Raíz del error cuadrático medio.
[4] Error absoluto porcentual medio.
[5] Error absoluto medio.

paso previo antes de afrontar un problema de aprendizaje supervisado. Dos de las formas más comunes de aprendizaje no supervisado son la reducción de dimensiones y el *clustering* (Chollet, 2018).

Sin embargo, existen otros tipos de aprendizaje que van más allá de la dicotomía del aprendizaje supervisado y no supervisado, siendo el aprendizaje auto supervisado y el aprendizaje reforzado dos grandes ejemplos. En el caso del aprendizaje auto supervisado, se sigue el mismo proceso que en el aprendizaje supervisado con la excepción de que los *outputs* con los que se entrena el modelo no han sido determinados por humanos, sino que se generan directamente desde los *inputs*. Respecto al aprendizaje reforzado, cabe recalcar que se trata de un tipo de aprendizaje que se encuentra en fase de investigación y a día de hoy no cuenta con grandes aplicaciones prácticas (Chollet, 2018). Se trata de un tipo de aprendizaje en el cual se generan estrategias de forma automática, por medio de la observación de experiencias y valoración de lo positivas o negativas que han sido dichas respuestas. En este sentido, es más cercano a la forma que tienen los humanos de razonar (Dimas, 2020).

Al igual que el *Machine Learning* es un subgrupo de la Inteligencia Artificial, el *Deep Learning* es un subgrupo del *Machine Learning*. Así las cosas, cabe plantearse qué aporta de especial el *Deep Learning* en relación con las técnicas de *Machine Learning* en sentido estricto.

Las técnicas de *Machine Learning* se mostraban insuficientes de cara a procesar los datos en su forma original –o forma bruta. Es decir, era necesario transformar esos datos para convertirlos en formas que el modelo pueda comprender, por medio de extracción o creación de nuevas *features*. Por el contrario, en el *Deep Learning*, al estar fundado en modelos que se sustentan en el aprendizaje basado en la representación –*Representation Learning*–, se pueden usar los datos en su forma original (LeCun, Bengio, & Hinton, 2015). Aunque desarrollaremos este concepto más adelante, las técnicas de *Deep Learning* usan modelos no lineales que permiten transformar los datos originales y abstraerlos en distintas capas de cara a la toma de decisiones. Esto será clave cuando nos enfrentemos a problemas cuya solución no sea lineal. Respecto a los datos originales, las técnicas de *Deep Learning* requieren una cantidad de datos y potencia de cómputo muy

superior a aquella requerida para las técnicas convencionales de *Machine Learning* (Chauhan & Singh, 2018).

El *Deep Learning* cuenta con propiedades que lo hacen muy atractivo como herramienta de *Machine Learning.* En concreto, su simplicidad –no hay necesidad de realizar *feature engineering*–, su escalabilidad y su versatilidad y reusabilidad –este tipo de modelos se pueden entrenar con datos adicionales sin la necesidad de empezar de cero (Chollet, 2018). En definitiva, el *Deep Learning* es una forma sofisticada de *Machine Learning* basada en las redes neuronales que nos permite realizar tareas mucho más complicadas, como puede ser la traducción automática o el reconocimiento facial.

Una vez analizadas las diferencias entre *Machine Learning* y *Deep Learning*, es interesante recalcar las disparidades entre los conceptos de *Shallow Learning* y *Deep Learning*. Ambos son dos subconjuntos del *Machine Learning*. El *Shallow Learning* serían todas aquellas técnicas convencionales que no pertenecen al campo del *Deep Learning.* Entonces, ¿qué caracteriza al *Deep Learning?* Recordemos que estas técnicas están basadas en las redes neuronales artificiales. Pese a que abordaremos este concepto más adelante, es necesario remarcar que en este contexto la diferencia entre las técnicas de *Shallow* y *Deep Learning* estriba en el número de capas intermedias entre los *inputs* y los *outputs*. En el caso del *Deep Learning*, se usan una gran cantidad de capas ocultas que permiten al modelo aprender representaciones de datos con múltiples niveles de abstracción (Vázquez, 2017). No existe un estándar para determinar a partir de qué número de capas ocultas se puede considerar que una red es de *Deep Learning.* Sin embargo, la mayoría de los expertos opinan que se consideraría *Deep Learning* desde el momento en el que se implemente una segunda capa oculta. (Vázquez, 2017).

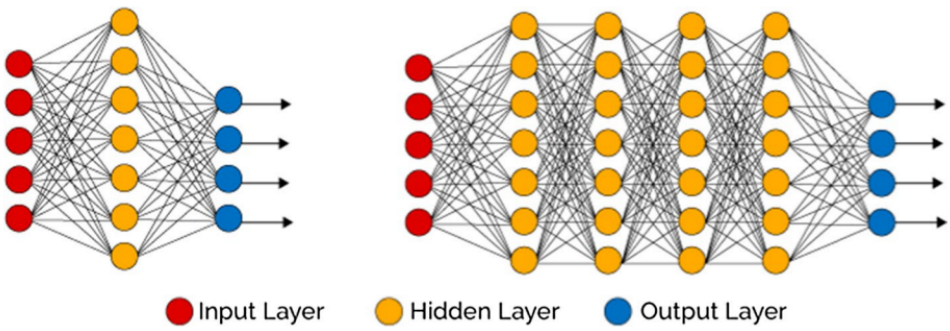


Figura 2
Shallow Learning v. Deep Learning en redes neuronales
Extraído de F.Vázquez, (2017) Deep Learning made easy with Deep Cognition. Medium.com. https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351
Nota A la izquierda encontramos una red neuronal en la que se aplica *Shallow Learning*, mientras que en la derecha se aplica *Deep Learning*

2.1. Concepto

Antes de explorar las redes neuronales profundas, es preciso comprender el funcionamiento de las redes neuronales artificiales en su conjunto.

En primer lugar, el funcionamiento y desarrollo de las redes neuronales artificiales o ANN (*Artificial Neural Networks*) se construyó con el funcionamiento de las redes neuronales biológicas del ser humano en mente. La referencia al modelo de la mente humana es común en el mundo académico, a modo de ejemplo: "Las redes neuronales artificiales son redes interconectadas masivamente en paralelo de elementos simples y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico" (Matich, 2001). De esta forma, una ANN sería un sistema inspirado en el modelo biológico humano que consigue establecer una serie de relaciones entre los *inputs* y *outputs* del modelo. Contrariamente a los modelos tradicionales de computación, en las ANN se procesan los datos en paralelo –no de forma secuencial. El modelo aprende los patrones presentes en los datos y usa las estructuras aprendidas para enfrentarse al problema en cuestión con nuevos datos.

2.2. El perceptrón

Para comprender el funcionamiento de las ANN es necesario referirse al perceptrón. Dicho concepto fue concebido por primera vez a finales de la década de los cincuenta por Frank Rosenblatt en *The perceptron –A perceiving and recognizing automaton.* La idea principal del perceptrón es actuar como una unidad básica de procesamiento Es decir, el perceptrón es a las redes neuronales artificiales lo que la neurona es al cerebro.

Al igual que en una neurona humana nos encontramos las dentritas que reciben las señales para pasarlas al cuerpo de la célula –o soma–, que reúne y suma todas las señales de entrada, que se transmiten a otras neuronas gracias al axón, en el perceptrón nos encontraremos con ciertos *inputs* x_i, i=1,...,M y una serie de pesos *wi*, i=1,...,M. Después se realiza la suma ponderada por los pesos de los *inputs* y se pasa por una función de activación no lineal que nos dará como resultado un *output.* Además, es preciso añadir un término de sesgo –*bias*– *wo* , que desplazará la función de activación a la izquierda o a la derecha.

$$\hat{y} = g\left(w_0 + \sum_{i=1}^{m} x_i \, w_i \right)$$

La relevancia de la función de activación en el perceptrón es máxima, puesto que transforma la suma ponderada de las entradas y los sesgos para introducir la no-linearidad en la red neuronal, lo cual es absolutamente necesario cuando nos enfrentamos a problemas que no son lineales, que son la mayoría (Soleimany & Amini, 2021).

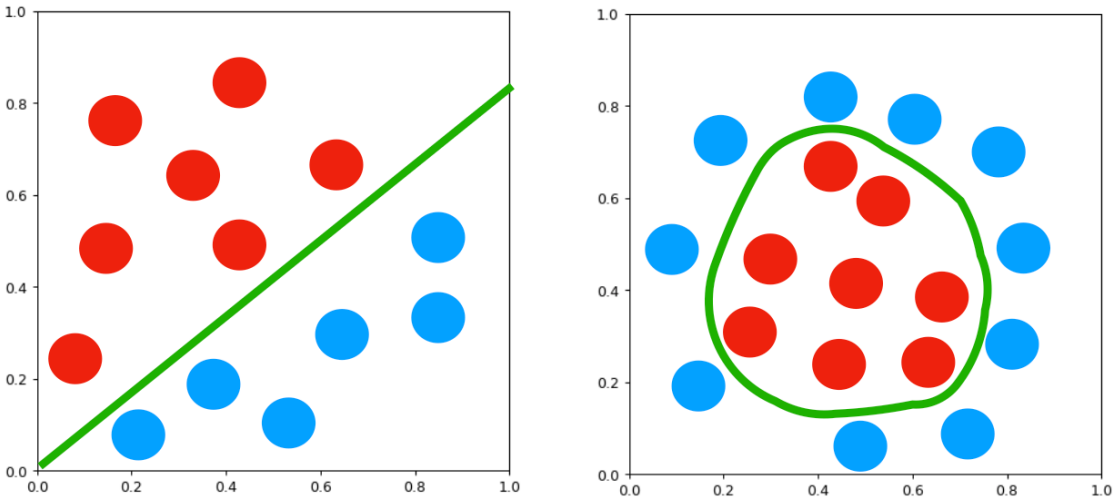
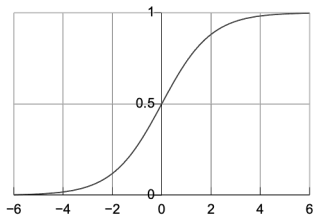
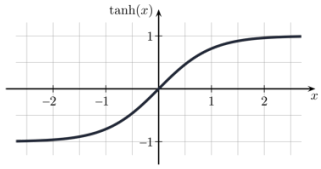
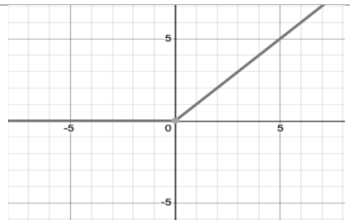


Figura 3
Problemas de clasificación linealmente separables y no linealmente separables
Elaboración propia

En la figura 3 nos encontramos a la izquierda con un problema de clasificación linealmente separable. Si aplicáramos una función de activación lineal no habría mayores problemas, pues esencialmente estaríamos llevando a cabo una regresión lineal. Por el contrario, a la derecha nos enfrentamos a un problema de clasificación que no es linealmente separable. Por lo tanto, será necesario acometer la tarea usando una función de activación no lineal.

Existen múltiples tipos de función de activación, siendo algunas de las más relevantes la función sigmoide –como LOGIT–, tangente hiperbólica o ReLU. Si bien las funciones sigmoides siempre han venido gozando de una gran popularidad, el uso de ReLU ha tomado mucha más fuerza gracias a su simplicidad y el hecho de que con las funciones sigmoides suelen surgir problemas de desvanecimiento del gradiente –entre otras características (Soleimany & Amini, 2021).

Función	Expresión matemática	Gráfica
Sigmoide	$g(z) = \frac{1}{1 + e^{-z}}$	
Tangente hiperbólica	$g(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	
ReLU	$g(z) = max\,(0, z)$	

Así las cosas, la representación gráfica del perceptrón sería la siguiente:

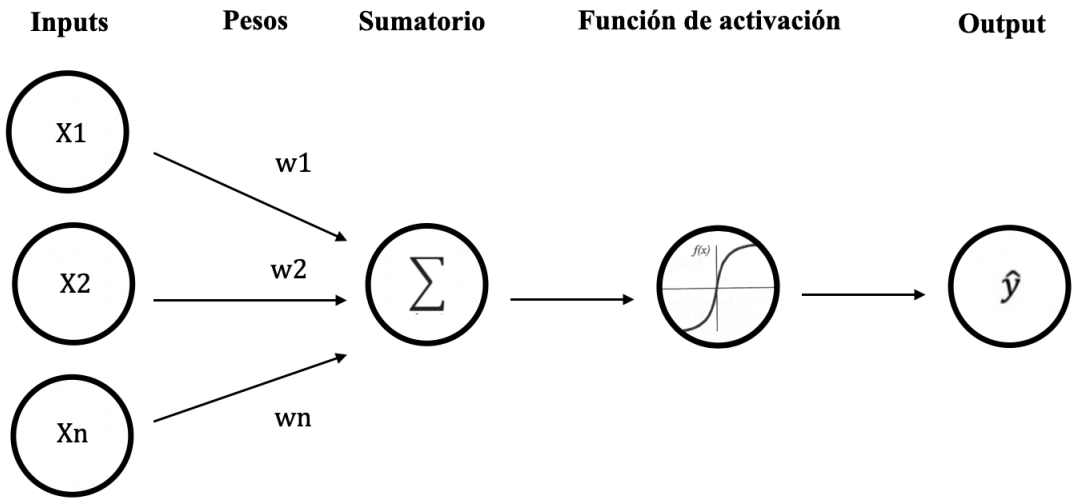


Figura 4
Esquema del perceptrón simple
Elaboración propia

2.3. El perceptrón multicapa – Redes neuronales multicapa

Una vez explorada la unidad básica de la red neuronal –perceptrón simple–, es necesario analizar el concepto de perceptrón multicapa. En este modelo encontraremos tres tipos distintos de capa: capa de entrada, capa o capas ocultas y la capa de salida. En la capa de entrada o *input layer* se recibe la información a procesar, mientras que en la capa de salida se realiza la tarea final de clasificación o predicción, donde se obtiene el *output*. Entre medias encontraremos una o múltiples capas ocultas, que constituyen el motor computacional del perceptrón multicapa (Abirami & Chitra, 2020).

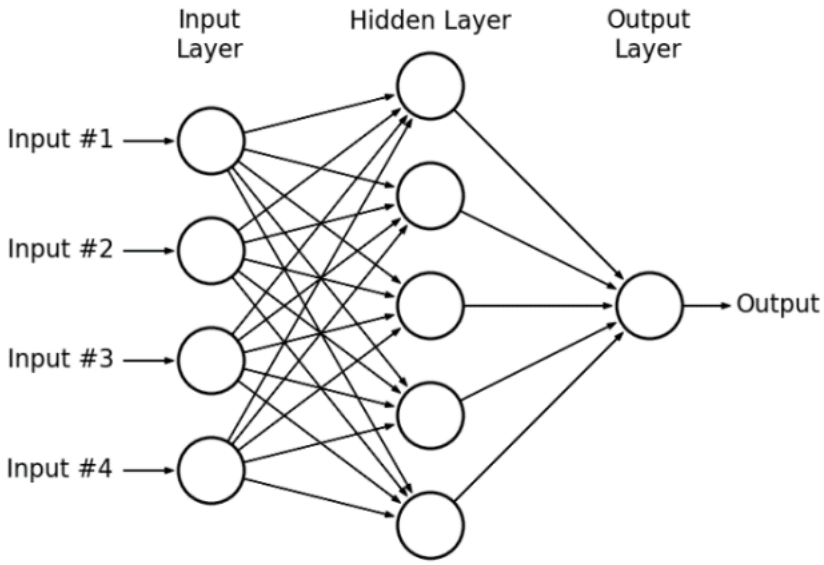


Figura 5
Esquema de perceptrón multicapa
Extraído de H. Mohamed Hassan, M.Zahran, A. Negm, O. Saavedra (2015). Assesment of artificial neural network for bathymetry estimation using resolution satellite imagery in shallow lakes: case study El Burullus Lake. International Water Technology Journal, IWTJ Vol. 5-No.4.

En la figura 5 encontramos un perceptrón multicapa con una capa oculta. Cabe mencionar que los perceptrones multicapa son *feed forward,* es decir, que no existen bucles; el resultado de la segunda capa oculta –en el caso de que existan varias– no volvería a pasar por la primera. Cuando nos encontremos con una red neuronal *feed forward* que permita lo contrario, estaremos ante redes neuronales recurrentes –concepto que desarrollaremos más adelante. Asimismo, la matriz de pesos de la primera capa y la segunda no es la misma sino que existen dos matrices diferenciadas. Además, en la figura 5 nos encontramos con capas densas –*dense layers* o *fully connected layers*–, pues todos los *inputs* están conectados de manera densa a la capa oculta, la cuál está conectada

totalmente a la capa de *output*. Es decir, el *input #1* está conectado a las neuronas que constituyen la capa oculta.

Cabe mencionar los avances que tuvieron lugar en alrededor de 2010 que permitieron dar el paso definitivo a las redes neuronales profundas. Antes, existían problemas de cara a la propagación del gradiente cuando había un número considerable de capas ocultas. En concreto, los avances que permitieron el entrenamiento de redes neuronales con más de diez capas ocultas fueron: mejores funciones de activación, mejores esquemas de inicialización de los pesos y mejores esquemas de optimización. Después, nuevos avances permitieron entrenar modelos con miles de capas ocultas, siendo la normalización de lotes un gran ejemplo (Chollet, 2018).

2.4. El proceso de aprendizaje en las redes neuronales artificiales

A través del proceso de entrenamiento de las redes neuronales artificiales, buscamos que la red pueda llevar a cabo tareas de clasificación o predicción con nuevos datos que no se hayan usado en la fase de entrenamiento y sin necesidad de intervención humana.

En el proceso de entrenamiento la red aprenderá a realizar la tarea modificando el valor de los pesos sinápticos, de tal forma que se minimice la pérdida. Los pesos iniciales suelen tener valores aleatorios, que cambian y se ajustan a medida que avanza el entrenamiento de la red neuronal (Cilimkovic, 2015). Para que el modelo pueda aprender, es necesario que el mismo sepa cuando realiza predicciones erróneas, de cara a poder corregirse a sí mismo. En concreto, la función de pérdida hace referencia a la diferencia entre el valor predicho por el modelo y el valor real, de tal forma que obtendremos una menor pérdida a medida que los *outputs* predichos y los *outputs* reales se parezcan más (Soleimany & Amini, 2021). En definitiva, buscamos encontrar aquellos pesos que nos permitan alcanzar la menor pérdida posible.

Existen varios tipos de funciones de pérdida, una de ellas siendo la raíz del error cuadrático medio, que es aplicable para modelos de regresión que tienen como *output* números reales continuos –por ejemplo, si quisiéramos predecir la nota que sacará un alumno en un examen. Otra función de pérdida bastante común es la entropía cruzada binaria, la cuál se aplica a modelos de clasificación binarios cuyo *output* es una

probabilidad entre 0 y 1. En concreto, la entropía cruzada binaria mediría como de alejadas estarían las distribuciones de probabilidad de los *ouputs* predichos y los reales. En este caso, la pérdida crece exponencialmente a medida que la predicción se aleja del resultado real (Seb, 2021) Finalmente, otro ejemplo común de función de pérdida sería la pérdida empírica, que mide la pérdida total en el dataset. Es decir, volviendo al ejemplo del alumno, la pérdida empírica sería la media de las pérdidas medidas para cada alumno –la diferencia entre la nota predicha para cada uno y la nota real (Soleimany & Amini, 2021).

Otra cuestión importante en el proceso de entrenamiento de las redes neuronales artificiales es el concepto de *backpropagation* o retropropagación del error. Se trata de un algoritmo que nos ayuda a realizar las correcciones necesarias a los pesos sinápticos que se establecen de forma aleatoria al principio del proceso de entrenamiento. En concreto, cada iteración del algoritmo se puede descomponer en cuatro fases principales (Cilimkovic, 2015):

1. Computación *feed forward:* Se obtienen los valores de las neuronas de las capas ocultas, usando dichos valores para determinar los valores propios de la capa de salida.
2. *Backpropagation* a la capa de salida: Se calcula el error en la capa de salida y se actualizan los pesos de la red desde la capa de salida hacia la capa oculta.
3. *Backpropagation* a la capa oculta: Se calcula el error en la capa oculta teniendo en cuenta los pesos actualizados del paso precedente y se actualizan los pesos entre la capa oculta y la capa de los *inputs*.
4. Actualización de los pesos sinápticos: Al final de la iteración, tendríamos todos los pesos actualizados en las conexiones con todas las capas.

El proceso se repite en tantas iteraciones como se desee, de cara a minimizar el error lo máximo posible. Además, es preciso mencionar que el algoritmo de *backpropagation* realiza su optimización aplicando el descenso del gradiente. En concreto, se trata de encontrar de manera analítica la combinación de valores de pesos sinápticos que den como resultado la menor pérdida posible (Chollet, 2018).

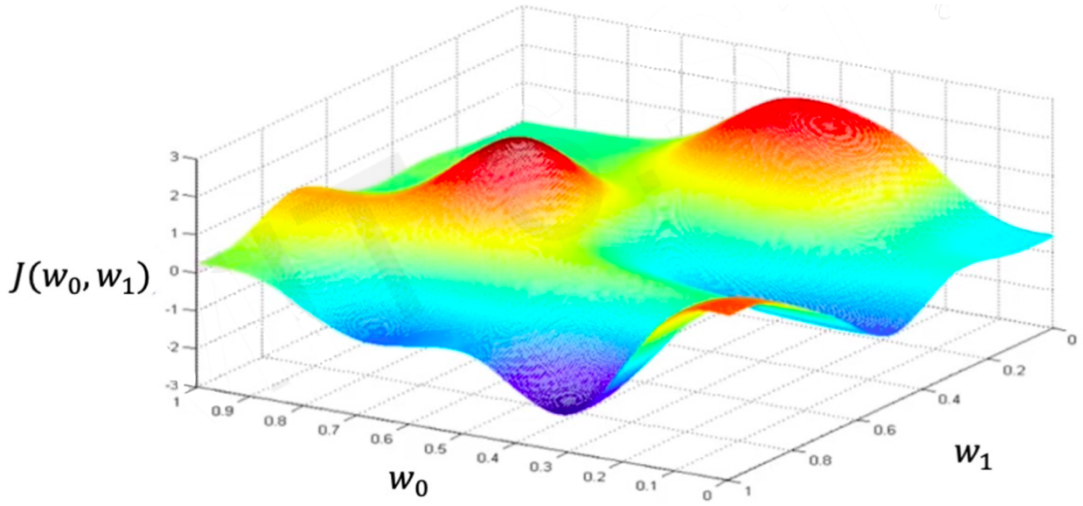


Figura 6
Función de pérdida con dos pesos
Extraído de Amini, Soleimany (2021) MIT Introduction to Deep Learning 6.S191: Lecture 1

En la figura 6 se puede apreciar la representación gráfica de una función de pérdida con dos pesos, $J(w_0, w_1)$. En primer lugar, se establecen los pesos de forma aleatoria. Después, se computa el gradiente $\frac{dJ(W)}{dW}$, que nos mostrará la dirección con mayor pendiente ascendente. Puesto que estamos buscando el mínimo de la función de pérdida, se modificará el valor de los pesos para tomar un paso en la dirección contraria a la que marque el gradiente. El proceso se repetirá hasta que alcancemos un mínimo local –puesto que los pesos se establecen inicialmente de forma aleatoria, no tenemos la certeza de alcanzar el mínimo global.

3. Redes Neuronales Profundas

Las redes neuronales profundas son herramientas clave en el desarrollo y aplicación de las técnicas de *Deep Learning*. Como mencionábamos anteriormente, las redes neuronales profundas estarán caracterizadas por la existencia de varias capas ocultas que permiten al modelo aprender una tarea llegando a varios niveles de abstracción (Vázquez, 2017). Es decir, se apilan dichas capas con el objetivo de crear modelos cada vez más jerarquizados donde el *output* se computa por medio de profundizar más y más en la red.

Existen distintos tipos de redes neuronales que se pueden aplicar según la tarea que se desee abordar. Por ejemplo, no se usará el mismo tipo de red para tareas de traducción

que para tareas de reconocimiento facial. En concreto, nos centraremos en las redes neuronales recurrentes y convolucionales.

3.1 Redes neuronales convolucionales

Como comentábamos *supra*, el *Deep Learning* ha gozado de gran relevancia en la última década. Entre las razones que explican este gran impulso que han vivido este tipo de modelos destaca la posibilidad de acceder a mayores volúmenes de datos públicos etiquetados de alta calidad, a la vez que la computación paralela basada en la GPU[6], pues su arquitectura en paralelo hace posible resolver varias tareas de forma simultánea y reduce así de manera considerable el tiempo de entrenamiento en los modelos de *Deep Learning* (Voulodimos, Doulamis, Doulamis, & Protopapadakis, 2018).

Uno de los campos más interesantes en los que se han adentrado las técnicas de aprendizaje profundo es el mundo del reconocimiento de imágenes. Los nuevos avances desarrollados en este contexto se han aplicado a infinidad de industrias. En este sentido podemos citar la robótica, la medicina, el reconocimiento facial, los coches de conducción autónoma y la accesibilidad para personas con discapacidades visuales. Por ejemplo, se han llegado a usar modelos de *Deep Learning* para analizar mamografías y detectar en ellas la presencia de tumores. Todas las aplicaciones citadas cuentan con un elemento común: el uso de redes neuronales convolucionales o CNN.

Las CNN encuentran sus primeros esbozos en *Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex* de Hubel y Wiesel, donde se comenzaron a desarrollar redes basadas en la arquitectura presente en el córtex visual animal -en concreto, de los gatos. Después, en la década de los ochenta serán especialmente relevantes los modelos computacionales fundados en conectividades locales entre neuronas y transformaciones de la imagen organizadas de forma jerárquica, en concreto el Neocognitron (Fukushima, 1980). Destaca asimismo el trabajo que Yann LeCun desarrolló desde Bell Labs, donde consiguió aunar las ideas del algoritmo de *backpropagation* y las incipientes CNN para enfrentarse a tareas de clasificación de dígitos escritos a mano (Chollet, 2018).

[6] Graphics Processing Unit

El objetivo que persiguen los modelos basados en CNN puede ser muy variado, pero todos ellos guardan relación con la meta de comprender a través de imágenes lo que está presente en ellas, dónde se encuentran los objetos que la componen, qué acciones están teniendo lugar y anticipar lo que vaya a ocurrir (Soleimany & Amini, 2021). Este tipo de tecnología se puede usar tanto para tareas de regresión como de clasificación, aunque nos centraremos en éstas últimas.

Cuando nos enfrentamos a un problema de clasificación de imágenes, extraer los *features* de forma manual no resulta práctico, pues estaríamos construyendo un sistema que sería vulnerable con respecto a las distintas variaciones que pueden presentar distintas imágenes de un mismo objeto (Soleimany & Amini, 2021). Por ejemplo, si queremos construir un modelo que detecte mesas, debemos tener en cuenta que existen mesas con un número variable de patas, que existen mesas redondas o cuadradas, que la iluminación en las imágenes no será homogénea y que cada foto de cada mesa estará hecha desde un ángulo distinto, entre otras cuestiones. Por ello, buscamos desarrollar un modelo que nos permita detectar y extraer dichas características o *features* de manera automática.

Es aquí donde entran en juego las CNN. Éstas cuentan con una multiplicidad de capas ocultas –todas ellas jerarquizadas–, de forma que las primeras capas de la red captarán las formas más simples –como bordes o machas oscuras–, mientras que las capas subsiguientes se especializan más y más a medida que son más profundas y van captando formas cada vez más complejas hasta poder al fin determinar el objeto en cuestión -en nuestro ejemplo, una mesa (Moreno, 2019).

Respecto a las particularidades de la arquitectura de las CNN, es preciso mencionar que muchas de las características que las diferencian de las ANN al uso provienen del hecho de que en las CNN los *inputs* serán imágenes, por lo que es preciso acomodar las características de la red a este tipo de datos (O'Shea & Nash, 2015). El grado de complejidad computacional que supone usar imágenes como *inputs* es demasiado elevado como para poder acometer este tipo de tareas usando ANN normales. A modo de ejemplo, si utilizáramos como *input* una imagen de 224x224 píxeles en blanco y negro tendríamos que tener más de 50.000 pesos conectados a la primera capa oculta.

Así las cosas, las CNN contarán con tres tipos distintos de capas, a saber: convolucionales, capas de *pooling* o de agrupación y *fully connected layers* o capas totalmente conectadas. Además, las neuronas de la red están organizadas en tres dimensiones: altura, anchura y profundidad -o número de canales.

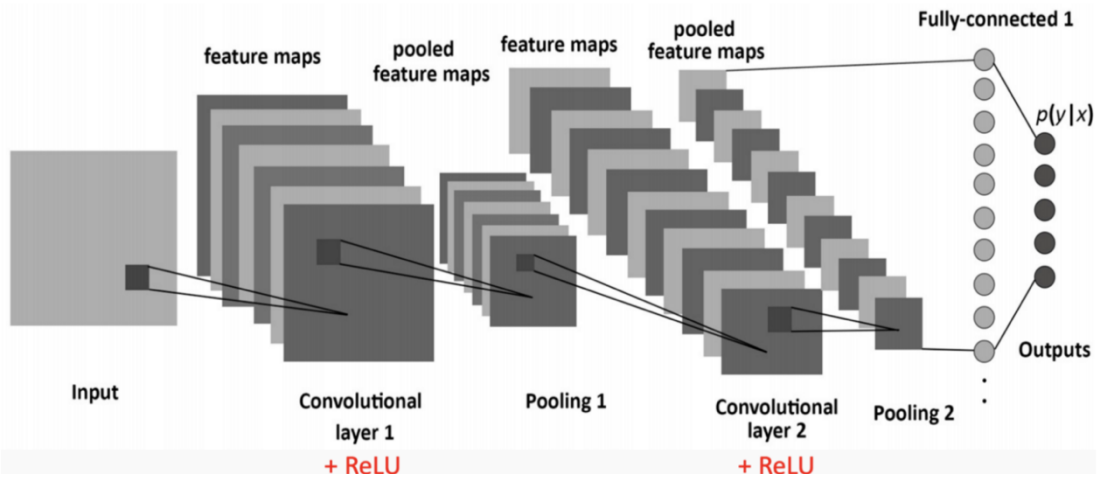


Figura 7
Arquitectura de red neuronal convolucional con dos capas convolucionales, dos capas de pooling y una fully connected layer
Extraido de Stewart (2019) Simple Introduction to Convolutional Neural Networks
https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac

A continuación, analizaremos cada tipo de capa presente en las CNN.

3.1.1 Capas convolucionales

Nos encontraremos con estas capas después de la capa de *input*, y éstas juegan un papel fundamental en el funcionamiento de este tipo de redes. El objetivo es tratar de reducir el número de conexiones que se dan entre la imagen *input* que estamos dando al modelo y las neuronas que se encuentran en la capa oculta. Esto se consigue conectando secciones del *input* de una determinada dimensión a una de las neuronas de la capa siguiente. Es decir, cada neurona de la capa está ligada a una pequeña y determinada parte del *input*.

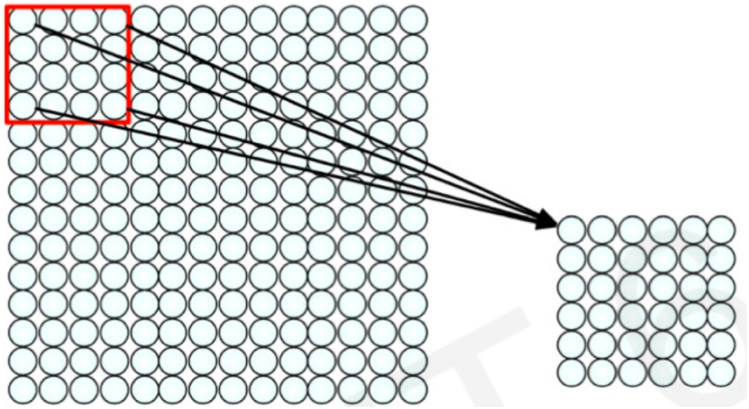


Figura 8
Conexión input-neurona en redes neuronales convolucionales
Extraído de Amini, Soleimany (2021) MIT Introduction to Deep Learning 6.S191: Lecture 3

En concreto, en este tipo de capas se aplicarán distintos filtros o *kernels* que extraerán características para crear un mapa de activación. Para cada valor en ese *kernel* se calcula el producto escalar, multiplicando los elementos que conforman el filtro en cada posición para posteriormente sumarlos. Es preciso mencionar que dicho filtro se desplazará a lo largo de la imagen, de la misma forma que haría una persona que lee a través de una lupa. Sin embargo, no será el único, pues se aplicarán múltiples filtros. De esta forma, se consigue analizar la totalidad de la imagen sin saturar las neuronas de la siguiente capa.

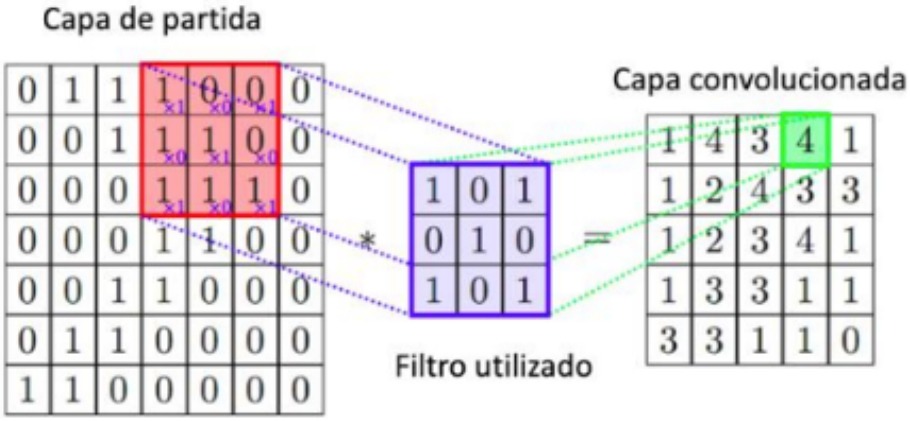


Figura 9
Funcionamiento de capa convolucional
Extraído de Calvo (2017) Red Neuronal Convolucional CNN https://www.diegocalvo.es/red-neuronal-convolucional/

Los tres hiperparámetros que optimizarán las capas convolucionales –y que por tanto, se pueden determinar según la decisión que se opte por tomar– son *depth*[7], *stride*[8] y el *zero padding* (O'Shea & Nash, 2015). El hiperparámetro *depth* se refiere al número de neuronas de la capa que se "conectará" a una sección determinada del *input*. En el caso de la *stride*, se refiere a cuánto se desplazará el filtro sobre la imagen desde una posición a la siguiente. Por su parte, el *zero padding* se refiere a la forma de procesar los bordes de la imagen *input.* Evidentemente, si escogemos valores bajos para dichos hiperparámetros, reduciremos la complejidad computacional del modelo –eso sí, a expensas la calidad del modelo de cara a reconocer patrones.

3.1.2 Capas de *pooling*

El objetivo principal de este tipo de capa es disminuir la dimensionalidad, reduciendo de manera considerable el tamaño de los antes mencionados mapas de activación. Esto no solo reduce la complejidad computacional de la red, sino que además nos ayuda a que el modelo no aprenda las posiciones exactas de las características en los *inputs* sino la estructura detrás de los mismos, haciéndolo más resistente a cambios de posición, iluminación o color de una imagen a otra (Brownlee, 2019).

Una vez más, estas capas funcionan a través de filtros, normalmente de 2x2, que se van moviendo a lo largo del mapa de activación hasta cubrirlo todo. Existen varios métodos de *pooling*, siendo los más comunes el *average pooling* y el *max pooling*. En ellos se toma la media de los valores de cada sección o el valor máximo del mismo según el caso.

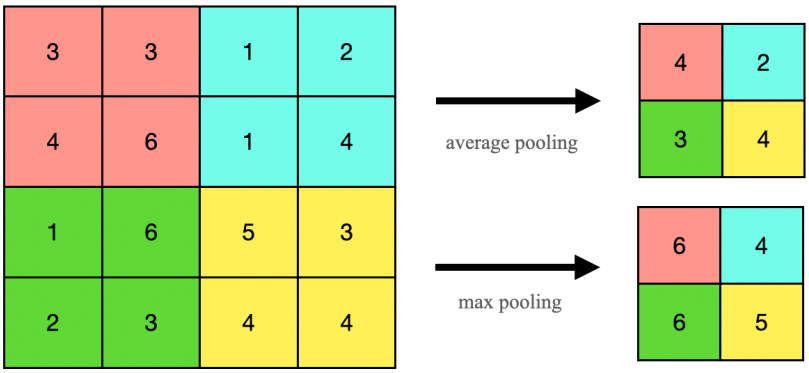


Figura 10
Comparación average pooling y max pooling
Elaboración propia

[7] Profundidad
[8] Zancada

3.1.3 Fully connected layers

La lógica detrás de este tipo de capa es completamente distinta a las anteriores y más similar a lo que hemos visto en el caso de las redes neuronales al uso. En ella, encontraremos neuronas que estarán totalmente conectadas a las neuronas de las dos capas colindantes (O'Shea & Nash, 2015).

Las *fully connected layers* o capas totalmente conectadas hacen las veces de "paso final" antes de la clasificación en sí, pues es en este punto del modelo donde toda la información recabada gracias a las divisiones por secciones que se han llevado a cabo y los filtros aplicados en las capas anteriores se agrupan. Dicho de otra forma, después de haber extraído las características de la imagen en el modelo, es preciso convertir todo lo que ha aprendido el modelo a un formato que nos permita llevar a cabo la tarea de clasificación.

Para ello, las *fully connected layers* convierten la información obtenida de las capas anteriores –mapas de características– en un único vector de una dimensión (Ignjatić, Bajic, Rikalovic, & Culibrk, 2018). Siguiendo el mismo proceso que en las redes neuronales normales explicadas en apartados anteriores, se aplicarán pesos y se dará una respuesta al problema de clasificación en la capa de *output*.

Por último, es preciso mencionar que en la última capa del modelo es bastante común el uso de la función de activación *softmax*[9], que nos permite obtener una distribución de probabilidad que se usará al clasificar, por ejemplo, si la imagen que hemos alimentado al modelo muestra o no una mesa.

En definitiva, las CNN son herramientas potentísimas para acometer problemas de clasificación de imágenes. En este tipo de modelos encontraremos capas que nos ayudan a extraer las características de la imagen –convolucionales y pooling– y capas como las *fully connected layers*, que nos ayudan a acometer la tarea de clasificación.

[9] Softmax (xi) = $\frac{exp(xi)}{\sum_j exp(xj)}$

3.2 Redes neuronales recurrentes

Cuando las personas leemos un libro, somos capaces de seguir la trama porque hemos conseguido retener la suficiente información de las páginas que hemos leído. Esta lógica se puede usar incluso a nivel de cada frase, pues conseguimos entenderlas gracias al hecho de que comprendemos y recordamos las palabras que hemos leído.

Este concepto de memoria es difícil de retener en las redes neuronales al uso, por lo que para este tipo de tareas debemos usar otro tipo de herramientas: Las redes neuronales recurrentes o RNN. Éstos son modelos extremadamente útiles cuando nos enfrentamos a tareas que presentan secuencialidad. A modo de ejemplo, se usa en procesamiento de lenguaje –pongamos el caso de un modelo que analiza un mensaje y predice el sentimiento que desprende– o en tareas de traducción. La clave en este tipo de redes es el hecho de que vamos a tener un componente temporal y secuencial.

Las principales razones por las que las RNN son especialmente interesantes para este tipo de cometidos son que éstas pueden procesar secuencias de longitud variable, además de su capacidad para detectar dependencias lejanas en el tiempo, conservar información sobre el orden –por ejemplo, de las palabras en una frase– y compartir parámetros a lo largo de la secuencia (Soleimany & Amini, 2021).

Las redes neuronales normales cuentan con unos *inputs*, una serie de capas ocultas –en el caso de las redes neuronales profundas– y un *output*. Las RNN son más sofisticadas, pues el *output* de un *time-step* se pasa al siguiente, creando un bucle que nos permitirá retener información pasada.

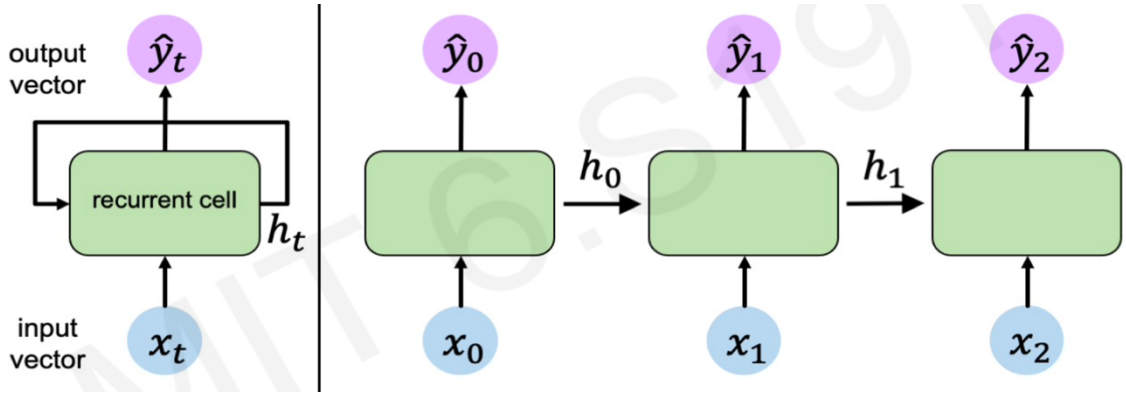


Figura 11
Red neuronal recurrente desarrollada
Extraído de Amini, Soleimany (2021) MIT Introduction to Deep Learning 6.S191: Lecture 2

De esta forma, los *outputs* del modelo no sólo serán una función de los *inputs* sino también de la memoria pasada de la siguiente forma: $\hat{y}_t = f(x_t, h_{t-1})$. Por supuesto, en este tipo de redes seguiremos teniendo pesos sinápticos, pero funcionan de forma distinta. Si en las ANN al uso –*feed forward*– tenemos distintos pesos asociados a cada neurona de cada capa, en las RNN se usa el mismo peso para cada capa de la red (IBM Cloud Education, 2020).

3.2.1 Backpropagation through time

Un aspecto importantísimo a tener en cuenta en este tipo de redes es el *backpropagation* a través del tiempo (BPTT). En los modelos *feed forward* se entrena el modelo tomando los resultados de cada iteración, calculando la función de pérdida y aplicando el algoritmo de *backpropagation*, que usa el descenso del gradiente para encontrar la combinación de pesos sinápticos que minimice lo más posible la pérdida. En las RNN debemos tener en cuenta el factor tiempo, puesto que la pérdida de un determinado *time-step* depende del *time-step* precedente (Torres, 2019). A través del BPTT, el *backpropagation* se lleva a cabo desde la pérdida total a cada *time-step* y después desde cada *time-step* hasta el principio. Así las cosas, el proceso que se sigue en las RNN para optimizar los pesos sinápticos es de una complejidad computacional nada desdeñable.

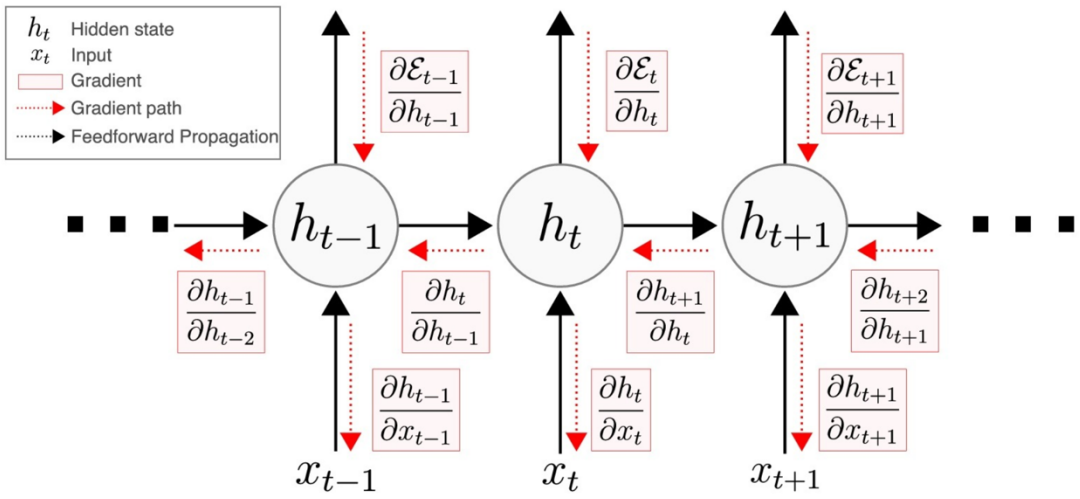


Figura 12
Funcionamiento del backpropagation through time (BPTT)
Extraído de T.P. Lillicrap y A. Santoro (2019). Backpropagation through time and the brain. Current Opinion in Neurobiology, Volumen 55, pág 83

Las RNN cuentan con dos grandes problemas a la hora de computar los gradientes para el *backpropagation*: el desvanecimiento del gradiente y la explosión del gradiente.

Si muchos de los valores asignados a los gradientes son mayores a 1, es muy posible que nos enfrentemos a un problema de explosión del gradiente. En esencia, este problema implica que se dé una relevancia desproporcionada a los pesos sinápticos y por tanto nos enfrentemos a problemas en el entrenamiento (Torres, 2019). En concreto, si los gradientes tienen valores demasiado altos, será extremadamente difícil llevar a cabo de forma exitosa el proceso de optimización. Para solventar el problema de explosión del gradiente, se truncarán los mismos para poder escalarlos.

Por el contrario, si muchos de los valores asignados a los gradientes son menores a 1 nos encontraremos con desvanecimiento del gradiente, lo cuál genera una serie de problemas. En concreto, al multiplicar números muy pequeños entre sí, los gradientes en *time-steps* que se encuentran en tiempos distantes serán prácticamente imperceptibles. Esto dificulta la propagación del error de nuestra función de pérdida a tiempos lejanos, lo que generaría sesgos en el modelo de tal forma que se tengan en cuenta de forma desproporcionada los patrones encontrados en tiempos más cercanos (Soleimany & Amini, 2021). Este problema se puede combatir desde distintos ángulos, pero la forma más eficiente es el uso de *gated cells.*

La idea es usar las *gated cells* como unidad básica del modelo, permitiendo controlar qué información se pasará al resto de *time-steps* a través de "puertas". En definitiva, se trata de conservar la información relevante y añadir la información nueva. Existen diversas formas de implementar las *gated cells*, destacando LSTM y GRU.

LSTM o *Long Short Term Memory* es un tipo específico de red neuronal recurrente propuesto por Hochreiter y Schmidhuber en 1997. Con el objetivo de poder conservar información relevante que pertenece a *time-steps* lejanos, se implementa una arquitectura en la que tendremos *memory blocks* que pueden contar con múltiples neuronas.

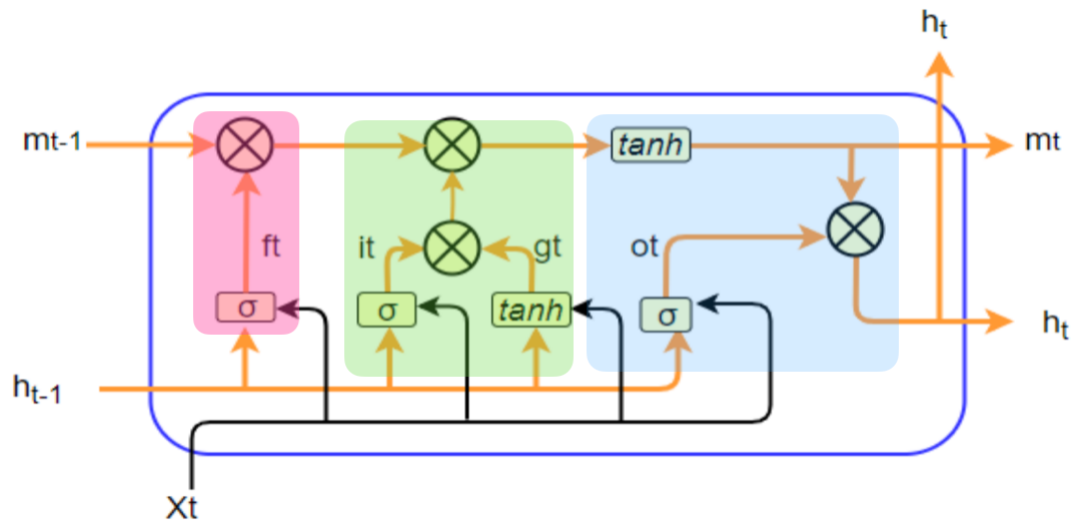


Figura 13
Estructura de LSTM
Extraído de L.H. Baniata, S. Park, S.B.Park (2018). A Multitask-Based Neural Machine Translation Model with Part-of-Speech Tags Integration for Arabic Dialects. Applied Sciences. 8. 2502.

A diferencia de las RNN al uso, cada neurona no cuenta únicamente con las entradas h_{t-1} y x_t –*input* en el instante *t* y memoria pasada– y una función de activación, sino que cuentan con tres puertas que nos permitirán no tener en cuenta la información irrelevante que venga de *time-steps* previos –zona roja en la figura 13–, almacenar aquella información nueva que el modelo estime relevante –zona verde– y determinar qué información se enviará al siguiente *time-step.*

Por último, es preciso destacar que las LSTM operan de forma similar a una función sigmoide[10] –rango de 0 a 1–, lo cuál permite incluirlas desde un punto de vista matemático en el *backpropagation* (Torres, 2019).

Por otro lado, el concepto de GRU o *Gated Recurrent Unit* se concibió en 2014, propuesto por Cho et al. en *On the Properties of Neural Machine translation: Encoder-Decoder Approaches.* Al igual que en LSTM, tendremos puertas que controlan el paso de información, pero en este caso no contaremos con *memory cells.* En concreto, existen dos tipos de puerta. En la puerta de actualización o *update gate*, se selecciona la información pasada relevante que se mantendrá en el modelo. En la puerta de *reset*, se decide qué parte de la información pasada podemos olvidar.

[10] $g(z) = \frac{1}{1+e^{-z}}$

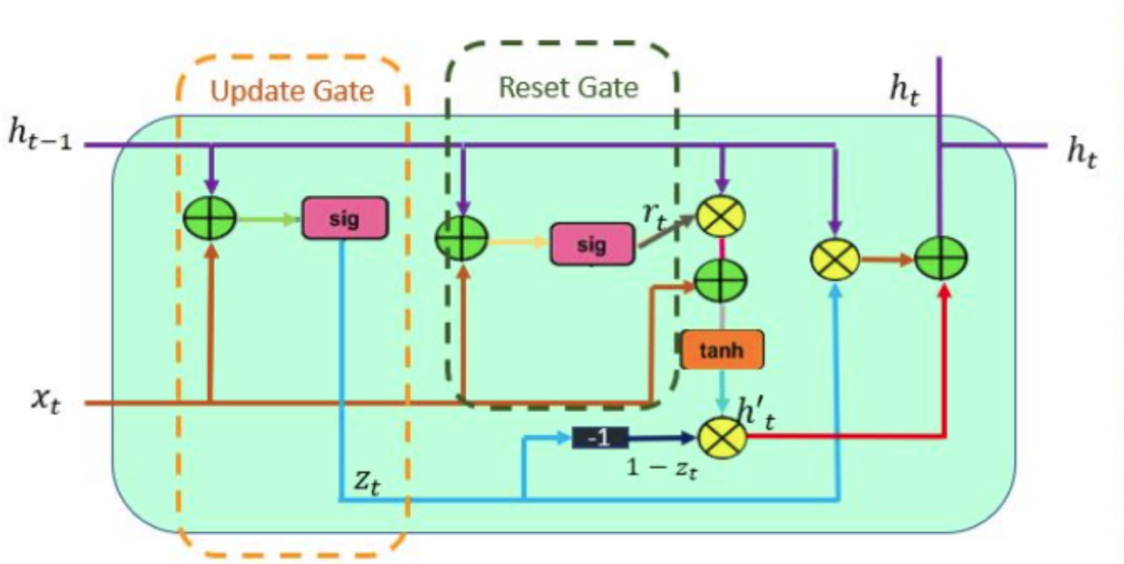


Figura 14
Estructura de GRU
Extraído de Lendave (2021) LSTMVs GRU in Recurrent Neural Network: A Comparative Study
https://analyticsindiamag.com/lstm-vs-gru-in-recurrent-neural-network-a-comparative-study/

Así las cosas, algunas de las mayores diferencias entre LSTM y GRU son el hecho de que en estas últimas nos encontraremos con dos puertas en vez de tres y que en las LSTM se controla la cantidad de información nueva que se almacena en la *memory cell* de forma independiente a aquella información de *time-steps* pasados que se decide descartar, mientras que en las GRU no se puede controlar de manera independiente la cantidad de información nueva que se almacenará (Chung, Gulcehre, Cho & Bengio, 2014).

3.2.2 Codificación del lenguaje

Las RNN cuentan con infinidad de posibles aplicaciones, pero resulta especialmente interesante su uso en tareas relacionadas con el lenguaje. Por ejemplo, modelos dedicados a la traducción o modelos que toman un texto y son capaces de seguir con la escritura, añadiendo frases nuevas.

Sin embargo, las redes neuronales no tienen la capacidad de procesar palabras directamente, sino que es preciso codificar el lenguaje de manera que podamos introducir en el modelo *inputs* numéricos que el modelo pueda comprender.

Para obtener un tensor numérico, es necesario vectorizar los datos, siendo las dos formas más comunes de vectorización el *One-hot Encoding* y el *Word Embedding* (Torres, 2019).

A través del *One-hot Encoding*, transformaremos nuestras frases en vectores binarios – contienen los valores 1 o 0 dependiendo de si la frase en cuestión cuenta con palabras de un vocabulario N. Se asignará un 0 para las palabras del vocabulario N que no estén presentes en las frases y un 1 si resulta que esa palabra del vocabulario está presente en la frase de *input*. Cabe destacar que no se aplica únicamente a frases, sino que se puede usar también para palabras sueltas o incluso caracteres.

	my	favourite	artist	is	Van Gogh
my	1	0	0	0	0
favourite	0	1	0	0	0
artist	0	0	1	0	0
is	0	0	0	1	0
Van Gogh	0	0	0	0	1

Figura 15
One-hot Encoding
Elaboración propia

Si bien ésta es una de las formas más comunes y sencillas de vectorización, la dimensionalidad es altísima, pues tendrá tantas dimensiones como palabras existan en el vocabulario (Chollet, 2018).

Por el contrario, con el *Word Embedding* la dimensionalidad es considerablemente más baja. Sigue siendo una técnica de vectorización, pero con una dispersión menor y con la capacidad de captar las relaciones semánticas entre las palabras. En concreto, es capaz de comprender el contexto en el que se da una palabra y ponerla en relación con otras. Es

decir, aquellas palabras que tengan un significado similar tendrán *word embeddings* parecidos (Lai, Liu, & He, 2016).

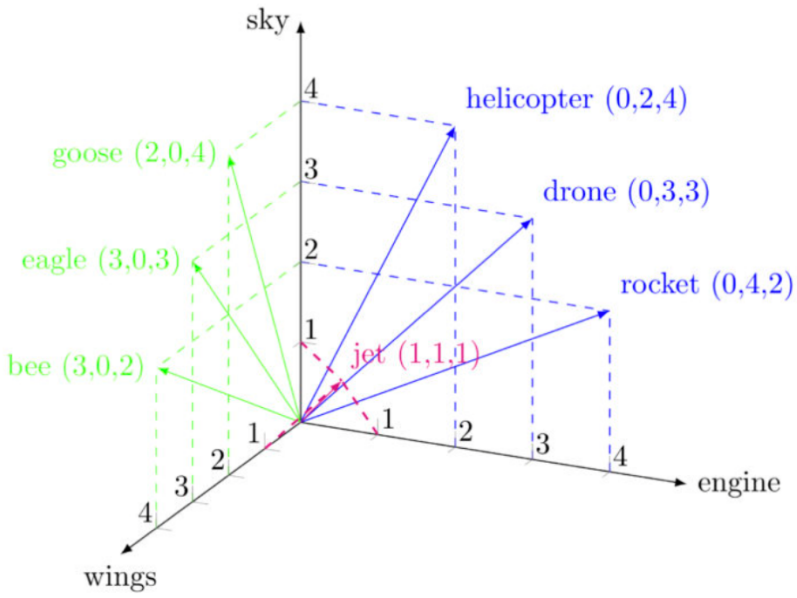


Figura 16
Espacio vectorial de word embeddings
Extraído de Desagulier (2018) Word embeddings: the (very) basics https://corpling.hypotheses.org/495

Como se observa en la figura 16, cada palabra se ve reflejada en el espacio vectorial y se encuentra cerca de palabras similares. Por ejemplo, la palabra *helicopter* es más cercana a *drone* que a *bee.*

En definitiva, la vectorización de palabras para poder nutrir al modelo de *inputs* numéricos que pueda comprender es una necesidad, y para ello tenemos dos vías diferentes. Las divergencias entre las dos técnicas de vectorización más comunes son esencialmente la alta dimensionalidad del *One-hot Encoding* y el hecho que éste no capta las relaciones semánticas entre las palabras.

3.2.3 Tipos de redes neuronales recurrentes

Como se ha incidido *supra*, las RNN se pueden utilizar en una enorme cantidad de contextos para solventar las tareas más variopintas. Esta variedad se verá reflejada en las distintas formas existentes de estructurar los *inputs* y los *outputs*. En concreto, nos centraremos en cuatro.

a) One to one

En este caso tendremos un *input* para un único *output*. Se trata de la forma más sencilla de red neuronal, pues realmente no hay un elemento de secuencialidad.

b) One to many

Contaremos con un *input* y múltiples *outputs*, es decir, el *input* en este caso no tiene un componente temporal. La estructura *one to many* se suele dar en tareas de descripción de imágenes (Soleimany & Amini, 2021). Por ejemplo, un modelo que toma como *input* una imagen de Kurt Cobain en un concierto de Nirvana y cuyo *output* sería la frase "una persona tocando la guitarra". Otro caso de modelos *one to many* serían los modelos de generación de canciones que son capaces de armar una canción siendo el *input* una única nota.

c) Many to one

En este caso, contaremos con múltiples inputs y un único *output*. Este tipo de estructura es especialmente común en modelos de análisis de sentimiento. Por ejemplo, un modelo que tomara la letra de una canción y fuera capaz de determinar si ésta es feliz o triste estaría usando la estructura *many to one*.

d) Many to many

Por último, en los modelos *many to many* contaremos con múltiples *inputs* y múltiples *outputs.* Un ejemplo paradigmático de este tipo de estructura serían los modelos de traducción, siendo los *inputs* las palabras de una frase en el idioma A y los *outputs* su traducción en el idioma B (Soleimany & Amini, 2021).

III. Redes neuronales profundas en el campo de la creación artística

Las redes neuronales profundas o *Deep Learning* se pueden aplicar a una infinidad de campos del conocimiento. A mi juicio, una de las aplicaciones más interesantes es el campo de la creación artística. Si bien tradicionalmente la tecnología y el arte se han concebido como disciplinas muy alejadas, actualmente se encuentran cada vez más y más cerca gracias al desarrollo de una serie de tecnologías –una de las más notables siendo el *Deep Learning*– que nos permiten cuestionarnos si el arte es una capacidad exclusivamente humana.

Las propuestas que unen arte e inteligencia artificial se podrían dividir en dos grandes categorías según su propósito: creación artística y análisis de arte existente (Cetinic & She, 2021). En este capítulo expondremos distintos ejemplos de la aplicación de las redes neuronales profundas en ambas categorías. En primer lugar, estudiaremos brevemente el potencial del *Deep Learning* aplicado al análisis de obras de arte. Después, exploraremos sus aplicaciones en la edición de imágenes, haciendo hincapié en el Neural Style Transfer, que nos permitirá aplicar el estilo de una imagen al contenido de otra –aunque también es aplicable a otras disciplinas que van más allá de las imágenes. Después, desarrollaremos la aplicación de las redes neuronales profundas en la generación de textos, que hacen posibles tareas tan interesantes como la generación de poemas o guiones de cine.

1. *Deep Learning* en el análisis de obras de arte

Antes de adentrarnos en la vertiente creativa del *Deep Learning*, es necesario hacer referencia a su potencial como herramienta para comprender y clasificar obras de arte.

El requisito previo al desarrollo de estas herramientas es realizar una labor de digitalización para poder contar con un repositorio que nos pueda servir como *input* para los modelos de clasificación y recuperación de obras. Una vez contamos con los *inputs* necesarios, es posible aplicar redes neuronales profundas –en concreto CNNs– para labores tan útiles como la clasificación automática de obras según el estilo o el artista. Es más, se han llegado a usar CNNs para reconocer determinados objetos, caras o motivos en las obras. Una buena muestra de ello el modelo de reconocimiento de personajes en

obras que desarrollaron Madhu et. Al en 2019, capaz de identificar, por ejemplo, si en una obra se está realizando una representación de la Virgen. Esto supone una herramienta muy interesante para los historiadores del arte, ya que la construcción de sistemas de recuperación puede ser muy útil para la organización y análisis de grandes colecciones de obras (Cetinic & She, 2021). Sin embargo, ésta no es la única forma en la que el *Deep Learning* puede ayudar a aquellos que se dedican al arte en su vertiente académica. Por ejemplo, Deng et. Al desarrollaron un modelo capaz de tomar las características de las obras de un autor –textura, colores, estilo– y medir la representatividad de una obra concreta. Es decir, el modelo sería capaz de evaluar lo acorde que es una obra al estilo general del artista (Deng, y otros, 2021). Esto supone de forma evidente una interesante herramienta para analizar obras de arte.

2. *Deep Learning* en la edición de imágenes:

Adentrándonos en el campo de la creación artística, el campo de las imágenes –ya sean obras pictóricas, fotografías o incluso vídeos– es sin duda uno de los contextos de mayor aplicabilidad de las técnicas de *Deep Learning*.

2.1 Google Deep Dream

Deep Dream es un algoritmo basado en CNNs desarrollado por investigadores de Google en 2015. En origen, buscaban entender el proceso de entrenamiento de este tipo de redes neuronales en cada capa. Sin embargo, acabaron obteniendo un algoritmo capaz de generar imágenes nuevas basadas en una imagen de referencia (McCaig & DiPaola & Gabora, 2016). El resultado concreto son imágenes psicodélicas en las que se exaltan los patrones de la imagen base.



Figura 17
Google Deep Dream
Extraído de Inceptionism: Going deeper into Neural Networks
https://photos.google.com/share/AF1QipPX0SCl7OzWilt9LnuQliattX4OUCj_8EP65_cTVnBmS1jnYgsGQAieQUc1VQWdgQ?pli=1&key=aVBxWjhwSzg2RjJWLWRuVFBBZEN1d205bUdEMnhB

En concreto, el algoritmo reconstruye la imagen en cada capa de forma que a medida que vamos avanzando en las capas, se captan patrones en las imágenes cada vez más complejos (E.Kurt, 2018). Esta técnica –denominada por sus creadores como *Inceptionism*– consiste básicamente en ordenarle a la red neuronal que profundice más en aquello que ve, de manera que si capta una imagen de una nube que se asemeja a un pájaro, transformará la imagen para hacerla más similar a un pájaro (Mordvintsev, Olah, & Tyka, 2015).

Así las cosas, Deep Dream es un claro ejemplo de lo conocido como creatividad transformacional, pues consigue transformar las dimensiones del espacio conceptual, generando estructuras nunca antes observadas (E.Kurt, 2018).

2.2 Neural Style Transfer

En 2015, los investigadores Gatys, Ecker y Bethge consiguieron desarrollar una revolucionaria técnica para la edición de imágenes: el *Neural Style Transfer*. Dicha técnica nos permite tomar dos imágenes, sacando el contenido de una –imagen *input*– y aplicando el estilo de otra –imagen de referencia. Esto permite, por ejemplo, tomar una fotografía cualquiera como *input* del modelo y obtener una foto con el mismo contenido que el *input* pero aplicando el estilo de Van Gogh o Munch.



Figura 18
Neural Style Transfer de Van Gogh y Munch
Extraído de L.A Gatys, A.S. Ecker, M.Bethge (2015). A Neural Algorithm of Artistic Style arXiv:1508.06576

Para conseguir resultados como los que se pueden apreciar en la figura 18, el *Neural Style Transfer* usa CNNs. Una de las cuestiones clave para desarrollar este tipo de tecnología fue descubrir que en las redes neuronales convolucionales es posible separar las representaciones de estilo y contenido. Recordemos que este tipo de redes aprenden información sobre las imágenes *input* de forma jerárquica, donde las primeras capas de la red toman las formas más simples. Para verdaderamente captar el contenido de la

imagen *input* de cara a posteriormente aplicarle un estilo distinto es necesario acudir a las capas más altas del modelo (Gatys, Ecker, & Bethge, 2015).

Para captar el estilo de la imagen de referencia, se usa en cada capa de la red lo que se conoce como matrices Gram, que captan las correlaciones entre los mapas de características en cada capa. En definitiva, para aprender y transferir el estilo se busca activar las neuronas de la red de forma que la imagen *output* y la imagen *input* se parezcan en el contenido y la imagen de referencia y la imagen *output* tengan las mismas características de estilo y textura en los mapas de activación (Baheti, 2021).

Como en cualquier algoritmo de *Deep Learning*, el concepto de función de pérdida es absolutamente imprescindible. El objetivo de esta función para el caso del *Neural Style Transfer* sería mantener el estilo de la imagen original y adoptar el estilo de la imagen de referencia (Chollet, 2018). En concreto, la función de pérdida seguiría la siguiente estructura:

L = distancia(estilo(imagen_referencia) – estilo(imagen_generada)) + distancia (contenido(imagen_input) – contenido(imagen_generada))

Cabe destacar que para calcular la distancia entre el estilo de la imagen de referencia y la imagen generada, se toma la distancia de las matrices Gram.

Así las cosas, la arquitectura básica de un modelo en el que se use el *Neural Style Transfer* para la edición de imágenes sería la siguiente:

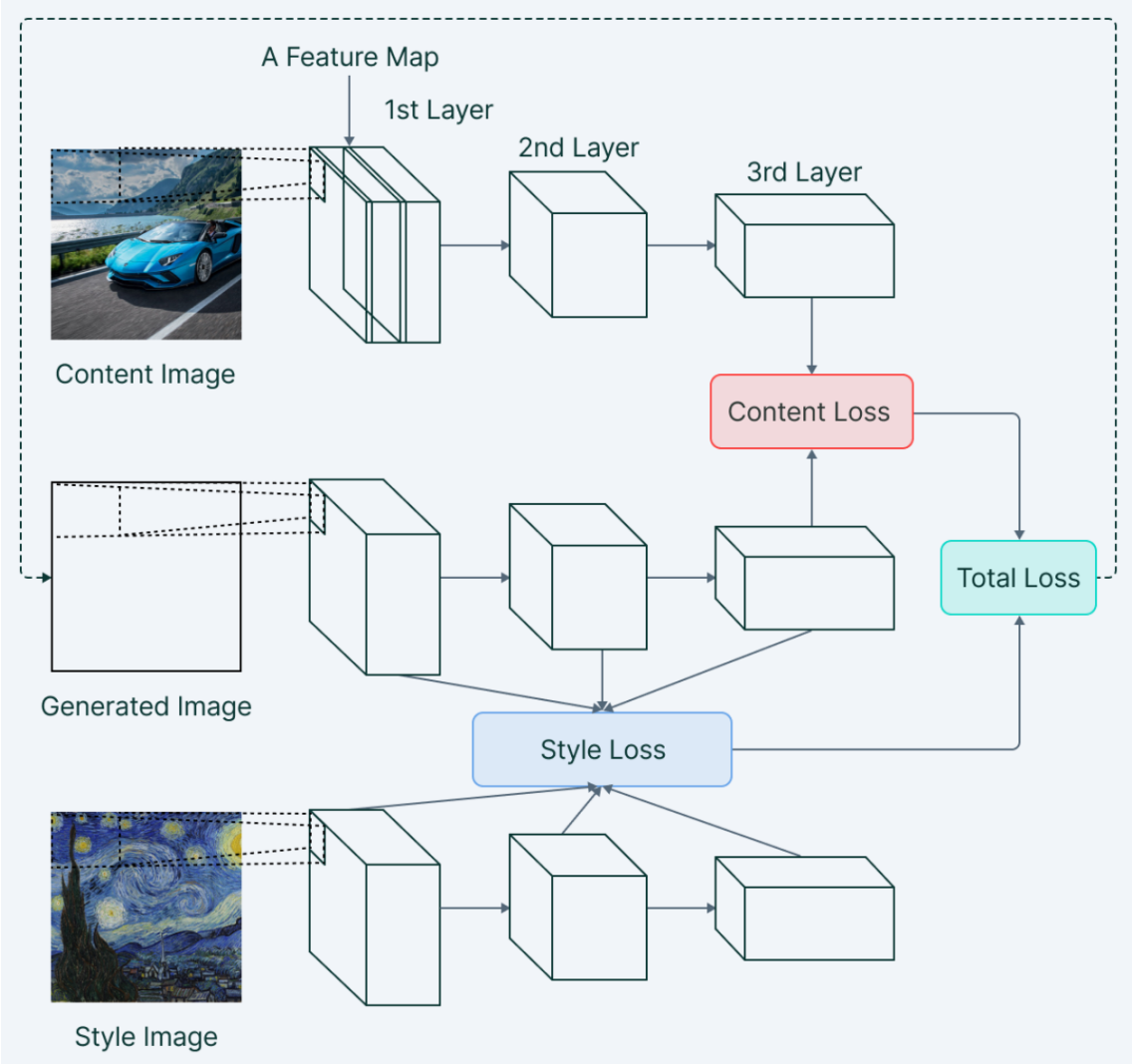


Figura 19
Arquitectura de un modelo de Neural Style Transfer
Extraído de P.Baheti (2021) Neural Style Transfer: Everything You Need to Know [Guide] https://www.v7labs.com/blog/neural-style-transfer

Es preciso mencionar que el si bien el *Neural Style Transfer* en su origen se aplicaba a la generación de imágenes, hoy en día se aplica a muchos otros campos del arte –no solamente a la fotografía y la pintura.

En el mundo de la música, se han llegado a aplicar modelos de *Neural Style Transfer* a espectrogramas para tomar la voz de un cantante y aplicarle el estilo sonoro de un violín, así como tomar el sonido de un arpa y aplicarle el estilo de un diapasón, abriendo la puerta a una nueva forma de sintetizar audio (Prateek & Smith, 2018). Sin embargo, es necesario destacar que la aplicación del *Neural Style Transfer* a la música trae consigo dificultades

añadidas en comparación con su aplicación a las artes plásticas. Con las imágenes es obvio pensar que los objetos y formas son el contenido mientras que las texturas y colores son el estilo. En la música no es tan intuitivo, ¿consideramos que la melodía es el contenido y la instrumentación el estilo o consideramos que la letra de una canción es el contenido y la melodía el estilo? Otra de las dificultades a la que se enfrenta la aplicación del *Neural Style Transfer* en la música es su componente temporal (Music Informatics Group, 2018).

Respecto al ámbito de la música y las artes plásticas es especialmente interesante el modelo que desarrollaron varios investigadores de la Universidad de Stanford en 2020, capaz de tomar música y convertirla en una imagen. Para ello, entrenaron una red neuronal profunda para que aprendiera a asociar fragmentos de música con la portada de su disco. En este sentido las imágenes con las que se entrenó el modelo fueron un elemento clave para que éste funcionara, optando por seleccionar las portadas de los discos por la principal razón de que, en general, las portadas se diseñan de forma que sean estilísticamente afines a la música que contienen, así como acordes a las emociones y al contenido de la música. De esta forma, se desarrolla un sistema que sin duda sería una herramienta muy interesante para los músicos, especialmente para puestas en escena en directos (Odlen, Verma, Basica, & Kivelson, 2020). En la misma línea, se han llegado a desarrollar modelos que modifican imágenes aplicando el estilo de una determinada pieza musical. En este caso, para relacionar música e imágenes en el dataset se tomó la década en la que tanto las imágenes como las piezas habían sido creadas. De esta manera se logró relacionar el contenido de las imágenes con el estilo musical de la época, consiguiendo que por ejemplo se pueda tomar una pieza musical de Debussy para transformar los colores de un cuadro de mitades del siglo XIX de forma que parezca una obra impresionista (Lee, Lin, Shih, Kuo, & Su, 2020).

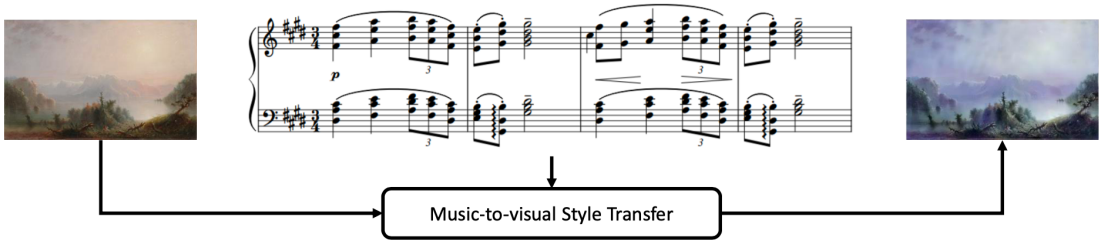


Figura 20
Neural Style Transfer de The Lake Her Lone Bosom Expands to the Sky (1850) de Jacob Miller con el estilo de Sarabande de Pour le piano, L. 95 (1901) de Claude Debussy. Extraído de Lee, Lin, Shih, Kuo, & Su (2020) Crossing You in Style: Cross-modal Style Transfer from Music to Visual Arts, MM '20: Proceedings of the 28th ACM International Conference on Multimedia pág 3219.

Por último, el *Neural Style Transfer* también es aplicable a archivos de vídeo. Es decir, se tomaría una imagen o un vídeo de referencia para aplicar su estilo al contenido de otro vídeo. Una de las mayores dificultades que esta tarea implica es el hecho de que si aplicamos la transferencia de estilo fotograma por fotograma de manera separada nos encontramos con discrepancias e inconsistencias temporales. Una de las formas de solventarlo sería tratar el vídeo *output* y aplicarle un post-proceso, pero ello impediría usar la transferencia de estilo para vídeos en tiempo real. Sin embargo, desde Tencent AI Lab algunos investigadores han desarrollado una forma alternativa de aplicar la transferencia de estilo en tiempo real para vídeos. Para ello, se ha utilizado una función de pérdida "híbrida" que tenga en cuenta por un lado la preservación del contenido del vídeo *input* al aplicarle los colores y texturas del vídeo o imagen de referencia, y, por otro, que tenga en cuenta la pérdida temporal de forma que se preserve la consistencia temporal entre distintos fotogramas (Huang, y otros, 2017).

3. Uso de redes neuronales recurrentes en la creación de textos

Originalmente, el *Neural Style Transfer* se desarrolló con el tratamiento de imágenes en mente. Sin embargo, una de las aplicaciones más interesantes de el concepto de transferencia del estilo se encuentra en el tratamiento de la palabra escrita. Esta tarea es especialmente complicada por dos razones principales, la primera siendo la dificultad para separar estilo de contenido en un texto. La segunda razón –a la que también se enfrenta el *Neural Style Transfer* aplicado a imágenes– es la dificultad en la evaluación, pues es complicado establecer una "verdad de referencia" para determinar el grado de calidad de la transferencia de estilo (Fu, Tan, Peng, Zhao, & Yan, 2018). Asimismo, es necesario recalcar que a diferencia de la transferencia de estilo en imágenes, la secuencialidad que implican los textos hace que las RNN sean especialmente adecuadas para acometer la tarea. De esta forma se han desarrollado modelos muy interesantes, por ejemplo modelos en los que se toma una frase en inglés estándar y se transforma al estilo de Shakespeare (Jhamtani, Gangal, Hovy, & Nyberg, 2017).

Dentro de las aplicaciones del *Deep Learning* en el mundo de la creación de palabra escrita, destacan los modelos que se han desarrollado para la generación de poesía. Investigadores de la Universidad de Southern California y de la Universidad de Washington desarrollaron en 2016 un modelo que era capaz de generar poemas basados

en la temática que eligiera el usuario. El modelo cuenta con una amplia base de vocabulario del que se conoce en qué sílaba tiene el acento. Una vez el usuario elige una temática –por ejemplo, *boda*–, el modelo computa una gran cantidad de palabras relacionadas con dicha temática –*novia, pareja, ceremonia*, *amor*–, de las que selecciona duplas que rimen para situarlas al final de las estrofas. De esta forma, se consigue que la temática se mantenga a lo largo del poema. Después, combinaron técnicas de *Deep Learning* –RNN– con máquinas de estado finito para garantizar una cierta corrección formal a la par que coherencia en la creación de las estrofas (Ghazvininejad, Shi, Choi, & Knight, 2016).

Respecto a las aplicaciones del *Deep Learning* en la generación de poemas a partir de una palabra seleccionada por el usuario, se han desarrollado en los últimos años una cantidad nada desdeñable de modelos que generan cuartetos en chino, siguiendo el estilo de la poesía clásica china preeminente en la dinastía Tang. En este sentido, es especialmente novedoso el planteamiento de Zhang y otros en *Flexible and Creative Chinese Poetry Generation Using Neural Memory*, pues abordan la problemática de la falta de originalidad en textos generados por redes neuronales. Debido a su manera de operar, las redes neuronales se suelen centrar en aquellos patrones que se repiten con frecuencia y acaba descartando aquellos que se presentan con menor asiduidad. Esto hace que se creen frases con sentido gramatical, pero se pierda una gran parte del elemento innovador para expresar sentimientos que tan importante es en la creación poética. Para solventar este problema, crearon una red neuronal de memoria aumentada capaz de recordar también patrones poco comunes, de la misma forma que un poeta se inspira –con frecuencia de manera subconsciente– en los poemas que ha leído a lo largo de su vida. Si bien es cierto que no estaríamos hablando de una genuina innovación, este modelo ayuda a crear una ilusión de novedad que otros no llegaron a alcanzar (Zhang, y otros, 2017). Sin embargo, el hecho de que no sea innovador –por basarse en ideas existentes– no significa que no se pueda considerar creativo. En concreto, se trataría de creatividad *combinacional*, cuyo valor reside en dar combinaciones nuevas de conceptos ya existentes (E.Kurt, 2018).

一山自有无人语，
Nobody speaking in the mountain,
不是青云入水边。
Also no green cloud stepping into the river.
莫把春风吹落叶，
Spring wind does not stir leaves,
花开绿树满江船。
But flowers blooming in trees and flying to boats.

Figura 21
Poema generado por ANN basado en la palabra "oneself" Extraído de Zhang, y otros, (2017) Flexible and Creative Chinese Poetry Generation Using Neural Memory.

Respecto a la generación de textos en prosa, la tarea es más complicada. Siendo computacionalmente más complejo, una de las razones que dificultan este tipo de tareas en comparación con los poemas es el hecho de que en la poesía hay más tolerancia a la ambigüedad. Es decir, los poemas se pueden interpretar por el lector de una forma mucho más libre que la prosa, por lo que ésta debe de ser más coherente para poder alcanzar un buen resultado (E.Kurt, 2018).

También es posible tomar las técnicas de generación de texto a través de RNN con el análisis de imágenes. En este sentido, destaca el trabajo de Jamie Kiros, investigadora de la universidad de Toronto que desarrolló la herramienta *neural-storyteller*. Dicha herramienta es capaz de tomar una imagen y crear un párrafo basado en la misma. Esto va más allá que una tarea de descripción de imágenes al uso, pues no sólo describe la imagen sino que crea una historia. En primer lugar, se entrena una red neuronal recurrente con distintos textos. Dependiendo del tipo de texto elegido para entrenar el modelo–novelas románticas, novela negra, clásicos… etc–, se obtendrán *outputs* acordes. En segundo lugar, se realiza un *embedding* de palabras e imágenes para mapearlas en un mismo espacio vectorial. Después, se usan vectores *Skip-thought* que tome la descripción asignada a la imagen y le aplique el estilo literario elegido (Kiros, 2015). A modo de aclaración, los vectores *Skip-thought* no son más que modelos de redes neuronales que se usan para aprender representaciones de frases longitud fija sin contar con datos etiquetados o aprendizaje supervisado.

Tomando el la herramienta *neural-storyteller* se pueden usar una infinidad de tipos de texto como *input* para entrenar al modelo y describir las imágenes de forma acorde. Por

ejemplo, se ha llegado a entrenar el modelo con letras de canciones de Taylor Swift para describir imágenes en el mismo estilo que la cantante americana.

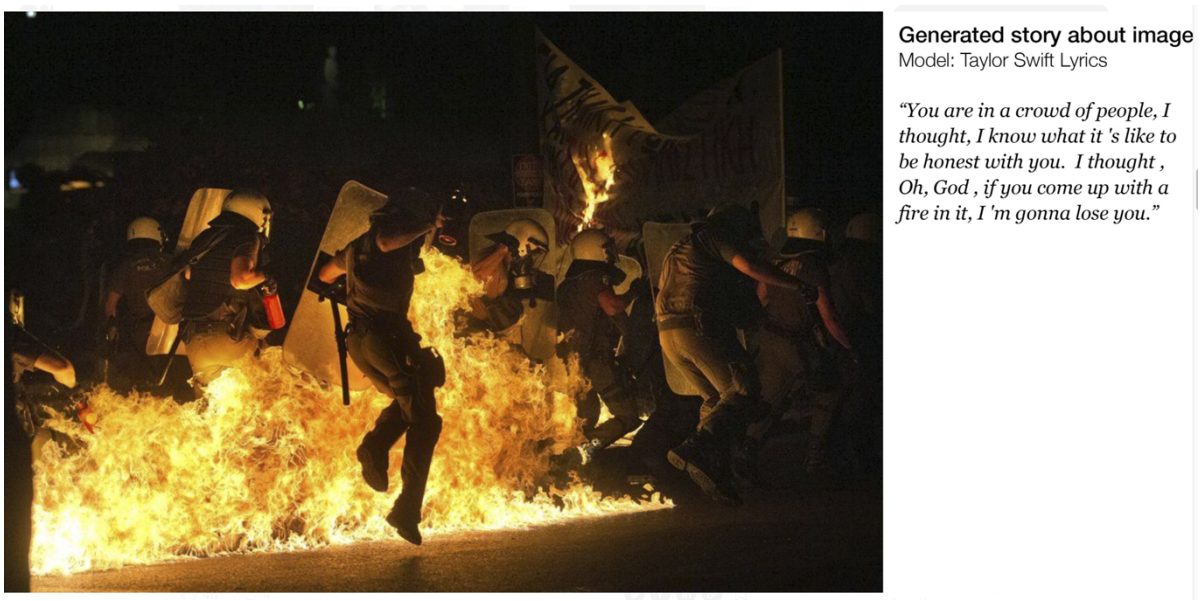


Figura 22
Descripción de imagen basada en letras de Taylor Swift
Extraído de samim (2015) Generating Stories about images. Medium.com.
https://medium.com/@samim/generating-stories-about-images-d163ba41e4ed

Uno de los ejemplos más curiosos de aplicación de RNN en la generación de textos es el cortometraje de ciencia ficción *Sungspring*. Dirigido por Oscar Sharp en 2016, el cortometraje causó especial revuelo en el festival de cine Sci-Fi London debido al hecho de que su guion fue redactado por una red neuronal recurrente LSTM. Para entrenar el modelo se usaron guiones de películas de ciencia ficción de los años 80 y 90, dando como resultado una historia protagonizada por tres personajes en lo que parece ser una estación espacial del futuro (Newitz, 2021). Sin embargo, es preciso mencionar que el guion está lleno de inconsistencias y frases extrañas que, si bien lo convierten en un experimento curioso, parece que todavía se está lejos de desarrollar una herramienta capaz de generar guiones que representen conversaciones realistas.

He picks up a light screen and fights the security force of the particles of a transmission on his face.

H
(continuing)
What do you mean?

C
(smiles)
I don't know anything about any of this.

H
(to Hauk, taking his eyes from his mouth)
Then what?

Figura 23
Extracto del guion de Sunspring
Extraído de https://www.docdroid.net/lCZ2fPA/sunspring-final-pdf

No contentos con el revuelo que suscitó *Sunspring*, el cineasta y su colaborador Ross Goodwin –artista y *data scientist*– decidieron participar en el mismo festival al año siguiente, entrenando seis modelos de redes neuronales recurrentes LSTM para obtener el guion de su segunda película *It's No Game*. Cada uno de los modelos fue entrenado con datos distintos. Por ejemplo, uno de ellos fue entrenado con diálogos de la época dorada de Hollywood, mientras que otro fue entrenado con diálogos de David Hasselhoff en la serie "El coche fantástico" –quien además protagonizó el cortometraje. Uno de los aspectos más interesantes de *It's No Game* es el hecho de que, a través de la utilización de distintos modelos, se consiguió crear distintas "voces narrativas" que se pudieron usar en cada escena según convenía (Newitz, 2017). Esto podría suponer en el futuro una interesante herramienta para los guionistas, de cara a asumir el difícil reto de desarrollar distintas voces narrativas a lo largo del guion.

IV. Implicaciones de negocio de las técnicas de *Deep Learning* aplicadas al arte

Como se ha podido comprobar, el *Deep Learning* cuenta con infinidad de posibles aplicaciones en el proceso creativo. Sin duda, se trata de técnicas que podrían servir de gran ayuda a los artistas. No obstante, es preciso tener en mente que el arte no sólo implica las piezas u obras en tanto que objeto artístico, sino que tiene grandes industrias detrás que mueven miles de millones de dólares cada año.

Con el objetivo de analizar las implicaciones que tiene el *Deep Learning* en las distintas industrias del arte, en este capítulo desarrollaremos en primer lugar las oportunidades de negocio surgidas en la industria de las artes plásticas. Después, nos adentraremos en la industria audiovisual, analizando las implicaciones del *Deep Learning* en las fases de preproducción y postproducción de proyectos. Finalmente, analizaremos las oportunidades que el *Deep Learning* supone para actores del mercado que no cuentan con muchos recursos –notablemente pequeños artistas.

1. *Deep Learning* en la industria de las artes plásticas

Las técnicas de *Deep Learning* cuentan con un futuro prometedor en el campo de las artes plásticas, especialmente en la pintura. Estas técnicas no suponen una mera herramienta que puedan usar los artistas para ayudar a crear su arte, sino que tiene claras implicaciones económicas tanto para los artistas como para los actores presentes en el mercado.

El *Deep Learning* en el arte ha trascendido el estatus de "herramienta" y se ha convertido en toda una oportunidad de negocio, siendo prueba el surgimiento de empresas cuya actividad se centra en la generación de obras de arte usando este tipo de técnicas. Un gran ejemplo es la compañía parisina Obvious. Esta empresa desarrolló su primera colección– *La Famille de Belamy*– en 2018, compuesta por 11 retratos generados por técnicas de *Deep Learning,* en concreto *Generative Adversarial Networks*[11]. La colección fue exitosa, siendo acogida con entusiasmo por los inversores y aficionados al arte. De hecho, la prestigiosa casa de subastas Christie's decidió subastar el cuadro *Edmond de Bellamy*. Pese a que el cuadro estaba estimado en unos 10.000$, el precio final fue de 432.500$

[11] Es un tipo de modelo que a su vez cuenta con dos submodelos: un modelo generador de nuevos ejemplos basándose en los *inputs* y un modelo discriminador que clasifica los ejemplos generados como reales o no, dependiendo de si considera que pertenecen al dominio o, por el contrario, falsos (Brownlee, 2019).

(*Manifesto Obvious Artificial Intelligence for Art*, 2018). Además, han conseguido exponer sus obras en los museos más importantes del mundo, notablemente el Hermitage. Sin embargo, Obvious no es el único ejemplo que demuestra que es posible sacar rédito económico de la creación de arte a través de redes neuronales profundas. AICAN es un algoritmo desarrollado por el profesor Ahmed Elgammal de la Universidad de Rutgers, que ha llegado a vender cuadros una cifra de 16.000$ *(Los algoritmos ya venden cuadros por miles de euros*, 2021).

Así las cosas, se puede afirmar que las técnicas de *Deep Learning* abren las puertas a un nuevo tipo de creaciones visuales que son muy interesantes para explotar en el mercado del arte desde el punto de vista del artista. Sin embargo, los artistas no son los únicos actores del mercado que se pueden beneficiar del surgimiento y desarrollo de este tipo de técnicas. Por ejemplo, los coleccionistas y casas de subastas se pueden beneficiar de herramientas basadas en las redes neuronales profundas para detectar falsificaciones. A este efecto encontramos empresas como Art Recognition, que usa CNN para determinar si una obra es un original, o por el contrario, se trata de un plagio (Schneider, 2020).

Además de suponer facilidades y oportunidades para los actuales actores del mercado del arte, el desarrollo de nuevas aplicaciones basadas en redes neuronales profundas tiene el potencial de permitir el acceso al mercado a inversores que no son especialistas. En el mercado de las artes plásticas ciertamente existen barreras de entrada debido a la complejidad y dificultad de conocer la calidad y el potencial especulativo de una obra sin ser un experto en la materia (*Can Artificial Intelligence change the art market?*, 2019). Con la idea de cambiar esta dinámica, la compañía americana Kellify ha desarrollado Naiada, una herramienta basada en el reconocimiento de imágenes capaz de predecir el valor de una pieza, así como futuras tendencias y posibles variaciones de valor (*Find out fine art hottest trends,* s.f.). De esta forma, el *Deep Learning* puede ayudar a reducir las barreras de entrada al mercado a nuevos inversores.

2. *Deep Learning* en la industria de la creación audiovisual

El campo de la creación audiovisual siempre ha estado ligado al progreso tecnológico, desde la invención de cinematógrafo a finales del s.XIX hasta el uso de efectos especiales avanzados de la actualidad, pasando por el fin del cine mudo y la inclusión del color en

los productos audiovisuales. Por ello, resulta especialmente interesante estudiar las propuestas de negocio que han surgido gracias a la aplicación de redes neuronales profundas.

Las técnicas de *Deep Learning* tienen cabida en todas las fases de la creación de películas y series, presentando nuevas oportunidades para los actores del mercado audiovisual. Desde guionistas hasta grandes estudios de cine, las redes neuronales profundas cuentan con el potencial para cambiar la forma en la que se aborda el proceso productivo de las creaciones audiovisuales, tanto a nivel de preproducción como de postproducción y promoción del producto.

Si bien uno podría pensar de forma intuitiva que la mayoría de las aplicaciones del *Deep Learning* para la industria que nos ocupa estarían orientadas al campo de los efectos especiales, existen propuestas de negocio muy interesantes a nivel de preproducción. Desde la selección de guiones, hasta el proceso de casting de actores pasando por la determinación del presupuesto del proyecto, las técnicas de *Deep Learning* supondrían interesante herramienta de la que se podrían beneficiar las productoras en especial (Raevskiy, 2020).

Uno de los mejores ejemplos que muestra que el análisis de guiones a través de técnicas de *Deep Learning* es una verdadera oportunidad de negocio es la empresa belga ScriptBook. El servicio que ofrecen es aplicar técnicas de *Deep Learning* y procesamiento de lenguaje para analizar guiones y recomendar a las productoras financiar o no un proyecto. La empresa ofrece un análisis en profundidad, determinando las emociones que transmite cada escena, si el guion pasa o no el test de Bechdel[12], prediciendo el rating que obtendrán en IMDb[13] y predicciones sobre la taquilla que obtendrá el proyecto, entre otras cuestiones (Raevskiy, 2020). Desde ScriptBook afirman que su algoritmo es tres veces mejor determinando los datos de taquilla que los humanos que trabajan en las productoras, además de asegurar que habrían sido capaces de detectar el fracaso de 22 de las películas de Sony Pictures con menor taquilla en los últimos tres años (Robitzski, 2018). Esta

[12] Test diseñado para analizar la brecha de género en el cine. Una película pasará el test de Bechdel si aparecen al menos dos personajes femeninos que mantienen una conversación entre sí cuya temática no gira alrededor de un hombre.
[13] Página web en la que se pueden encontrar detalles sobre películas, incluyendo el rating dado por los usuarios.

herramienta podría ser muy interesante para las productoras de cara a no sólo tomar decisiones de inversión basándose en la opinión humana de sus ejecutivos, sino tener también una base analítica.

Una de las propuestas de negocio más interesantes como herramienta de preproducción viene de la mano de Cinelytic. Esta empresa ha conseguido desarrollar una plataforma capaz de predecir la taquilla de una película basándose en los temas que trata la película y los actores que la protagonizan. De esta manera, los productores que usan la herramienta podrían comprobar cómo cambia el *forecast* dependiendo de si deciden contratar a un actor u otro (Vincent, 2019). Cabe destacar que la herramienta es capaz de hacer predicciones adaptadas a cada territorio. Por ejemplo, sería capaz de decirnos que cierto actor haría obtener una mejor taquilla en Europa mientras que otro mejoraría los datos en China. Además, Cinelytic es capaz de analizar posibles estrategias de distribución y financiación. Prueba del interés de las grandes productoras en este tipo de herramientas es el hecho de que en 2020 Cinelytic cerró un acuerdo de colaboración con Warner Bros para usar su herramienta de gestión de proyectos.

Así las cosas, se puede afirmar que las herramientas basadas *Deep Learning* tienen el potencial de convertirse en servicios esenciales para las productoras, puesto que conseguirían reducir el tiempo que sus ejecutivos dedican a evaluar las propuestas que les llegan. Además, la inmediatez que caracteriza a este tipo de herramientas las hace especialmente valiosas en contextos como los festivales de cine, donde los estudios cinematográficos deben pujar con extraordinaria rapidez por películas sin tener el tiempo de analizar a fondo las propuestas (Siegel, 2020).

Alejándonos del terreno de la preproducción, las técnicas de *Deep Learning* podrían convertirse en una herramienta muy interesante para la elaboración de trailers. De hecho, desde IBM se consiguió elaborar el tráiler de la película *Morgan* usando un modelo al que se le habían dado como *input* más de 100 trailers de películas de terror. El modelo fue capaz de seleccionar 6 minutos de la película, que después usó un editor para elaborar un tráiler de poco más de 2 minutos. El impacto que este tipo de tecnologías puede tener para las empresas cuyo modelo de negocio se basa en ser subcontratado para elaborar trailers es enorme, pues el proceso de 10 a 30 días que supone elaborar un tráiler se reduciría a 24h (Heathman, 2016).

Adicionalmente, las técnicas de *Deep Learning* tienen una gran cabida en la postproducción audiovisual. En este terreno, la industria ya no sólo busca artistas de efectos especiales con grandes dotes creativas, sino que también busca ingenieros con experiencia en *Deep Learning* que puedan ayudar a facilitar el trabajo de los primeros (Robitzski, 2018). Un gran ejemplo de ello es el trabajo que se llevó a cabo desde la empresa Digital Domain para la película *Avengers: Infinity War*. En este caso, se consiguió transformar al actor Josh Brolin en Thanos a través de técnicas de *Deep Learning* prácticamente a tiempo real, cuando es una tarea que normalmente lleva semanas (Robitzski, 2018). Claramente, esto tendrá implicaciones negativas para los artistas de efectos especiales, especialmente si se generaliza su uso fuera de la burbuja de los grandes estudios cinematográficos.

También en el mundo de la postproducción audiovisual nos encontramos con nuevas empresas que han decidido hacer de las técnicas de *Deep Learning* aplicadas a las imágenes su modelo de negocio. Este es el caso de Runway, una *startup* que ha creado una plataforma de edición de video que permite, por ejemplo, eliminar una taza de café que accidentalmente sale en el plano de una película de época (Vincent, 2019). Propuestas como estas permitirían acercar al gran público efectos y animaciones que anteriormente sólo podían alcanzar los grandes estudios, democratizando su acceso. Esto podría suponer toda una revolución en el sector de los efectos especiales.

En definitiva, las redes neuronales profundas podrían cambiar la manera en la que tradicionalmente se ha llevado a cabo la toma de decisiones en la industria del cine, así como servir de base para herramientas que podrían ser de utilidad para la gran mayoría de actores del mercado, aunque la automatización de ciertas tareas ciertamente puede tener consecuencias en términos de empleo.

3. *Deep Learning* en la industria musical

La industria musical ha sufrido grandes cambios en las últimas décadas. La época de los CDs y el monopolio de las grandes casas de discos ha terminado. Según el *Global Music Report* de la IFPI[14], el 62.1% de los ingresos mundiales procedentes de la música grabada

[14] International Federation of the Phonographic Industry

correspondieron a servicios de *streaming*. La digitalización de la industria es evidente, lo por lo que la aplicación de las técnicas de *Deep Learning* podrían ser especialmente interesantes en este campo. En concreto, nos centraremos en las herramientas que facilitan la labor de aquellos actores de la industria que no cuentan con grandes medios –especialmente a los pequeños artistas.

En primer lugar, la generación de música ayudándose de técnicas de *Deep Learning* se ha convertido ya en un modelo de negocio. Uno de los mejores ejemplos es la compañía Aiva Technologies. Esta compañía usa redes neuronales profundas con el objetivo de crear bandas sonoras para anuncios y documentales. La generación estaría automatizada, eligiendo en usuario una serie de parámetros como el estilo o la emoción a transmitir (Briot, 2021). Si este tipo de negocios se generaliza y la generación automática de música toma fuerza, se darían cambios en las formas de producción de la industria, así como pérdidas de puestos de trabajo (Weng & Chen, 2020). Sin embargo, son una opción muy interesante para otros productos culturales –por ejemplo, películas– que no cuentan con el presupuesto suficiente para contratar los servicios de un compositor.

Sin embargo, la creación musical no es el único ámbito de la industria en las que las técnicas de *Deep Learning* son de gran utilidad. La empresa barcelonesa BMAT las aplica al campo de la detección de música, constituyendo una herramienta muy interesante para los artistas. Vericast –una de las herramientas de BMAT–, es capaz de rastrear el uso que se le da a una pieza musical de cara a tener un control desde el punto de vista del copyright. La herramienta es incluso capaz de detectar música de fondo, lo cual ha venido siendo un quebradero de cabeza para las tecnologías de seguimiento de música (Sciforce, 2021).

Otro de los aspectos en los que el *Deep Learning* podría ser de gran ayuda en la industria –particularmente para los artistas pequeños– es en la masterización de audio. Grabar una canción con calidad no es tarea fácil, especialmente si no se cuentan con los medios necesarios para contratar a especialistas. Uno de los pasos ineludibles para conseguir esa calidad sonora es la masterización, pues se consigue equilibrar los elementos sonoros de la mezcla para optimizar la calidad de la reproducción, asegurando la cohesión del sonido en diversos formatos y entornos (Birtchnell, 2018). Empresas como LANDR han conseguido identificar la necesidad de facilitar el acceso a pequeños artistas y productores a servicios de masterización, usando para ello técnicas de *Deep Learning*. Propuestas

como esta claramente reducen las barreras de entrada al mercado a los pequeños artistas que buscan publicar canciones de calidad desde el punto de vista del audio sin tener que adquirir conocimientos técnicos de masterización ni sufragar los costes de contratar a un profesional.

En definitiva, las técnicas de *Deep Learning* ya se encuentran presentes en la industria musical, constituyendo herramientas muy interesantes que democratizan el acceso a la industria a actores del mercado que no cuentan con muchos recursos.

V. Reflexiones sobre el *Deep Learning* aplicado al arte

Llegados a este punto, resulta evidente el hecho de que el *Deep Learning* está en condiciones de generar grandes cambios no sólo a nivel industria, sino también en la forma que concebimos el arte.

En este capítulo se expondrán una serie de reflexiones acerca del *Deep Learning*, abordando la cuestión de si los modelos de *Deep Learning* llevan a cabo actividades verdaderamente creativas, así como la cuestión de si se puede considerar arte. Después, analizaremos las implicaciones de la inteligencia artificial en el arte en términos de derechos de autor. Finalmente, expondremos algunas de los problemas éticos que la aplicación de este tipo de técnicas puede suponer.

1. ¿Son los modelos de *Deep Learning* capaces de generar arte?

El desarrollo de técnicas de inteligencia artificial aplicadas al arte cuestiona el monopolio humano sobre los conceptos de creatividad y arte. Si bien no hay una respuesta unívoca a la cuestión de si los modelos de *Deep Learning* son capaces de generar verdadero arte, resulta interesante explorar las distintas reflexiones que se han llevado a cabo sobre la materia.

La creatividad y el arte se han considerado a lo largo de la historia como fenómenos propios del ser humano. Sin embargo, se trata de un concepto misterioso y ambiguo que varía según cada cultura. Partiendo de la concepción estándar de creatividad entre los expertos, uno de los aspectos clave en la creatividad es la novedad. En concreto, Stein sostiene que los procesos creativos deben necesariamente culminar en una obra nueva que sea considerada útil por un grupo de personas en algún momento del tiempo. También la intencionalidad sería un elemento intrínseco a la creatividad (Klausen, 2010). En este sentido, Amabile indicaba que para que una obra sea creativa es necesario que sea una forma nueva y valiosa de lidiar con la tarea que se trate, así como que la tarea sea heurística y no algorítmica (Klausen, 2010).

Ciertamente, algunas concepciones sobre la creatividad chocan con la idea de arte generado por *Deep Learning*. ¿Son capaces los modelos de *Deep Learning* de aportar

novedad? Es cierto que un modelo nos dará resultados basados en los *inputs* que seleccionemos, y no puede generar ideas nunca antes vistas. Sin embargo, no existe un único tipo de creatividad. En este sentido, Sawyer habla de creatividad combinacional. Es decir, realizar combinaciones nuevas de conceptos existentes se puede considerar como una actividad creativa (E.Kurt, 2018).

Respecto a la intencionalidad, Mikalonytė y Kneer argumentan que existen 5 elementos propios de la intencionalidad, y la inteligencia artificial cumpliría con todos ellos: (i) deseo de obtener un resultado, (ii) una creencia acerca de la acción de lleva a dicho resultado, (iii) la intención de realizar la acción), (iv) conciencia del cumplimiento de la intención mientras se realiza la acción y (v) habilidad para realizar la acción (Mikalonytė & Kneer, 2021). Aunque la inteligencia artificial no cumpliera con todos los requisitos, es indudable que existe intencionalidad por parte del programador. Los modelos de *Deep Learning* serían un medio de expresión artística, pues el programador elige los datos de *input*, programa el modelo, hace un tratamiento del *output* y cuenta con la intencionalidad de crear algo novedoso. Es preciso hacer énfasis en el hecho de que no estamos ante una simple herramienta como pudiera ser un pincel o un instrumento musical, sino ante un medio de creación. Las técnicas de *Deep Learning* van más allá puesto que es capaz de cambiar a medida que el modelo se entrena y tomar decisiones basadas en experiencia pasada (Mazzone & Elgammal, 2019).

Los detractores de la idea de que se pueda crear arte a través de inteligencia artificial argumentan que las máquinas no son conscientes de lo que hacen, que no tienen preferencias artísticas, y que las obras de arte son la expresión de experiencias puramente humanas (E.Kurt, 2018). Muchos de estos argumentos centran su definición del arte y la creatividad en la figura del artista. Sin embargo, la concepción de arte como proyecto personalista que expresa las emociones de un ser humano concreto es moderna, datando del s.XIX y XX en Europa Occidental (Mazzone & Elgammal, 2019). Por lo tanto, no existe una definición atemporal y universal de arte. Es evidente que no se puede replicar la psique y expresión del artista desde el *Deep Learning*, pero eso no significa que las obras fruto de su aplicación no puedan ser consideradas obras de arte.

En vez de construir el concepto del arte sobre la figura del artista, es posible construirlo sobre la obra en sí y la audiencia que la admira. Según Sawyer, cuando una obra es

admirada, empieza a formar parte de la cultura humana y constituye un valor creativo desde el punto de vista sociocultural (E.Kurt, 2018). Es decir, pasaríamos a definir el arte según su capacidad para ser admirada por un ser humano y transmitirle emociones. En esta concepción del arte cabe sin duda aquel generado por *Deep Learning*, especialmente si tenemos en cuenta que se ha llegado al punto de conseguir que obras de arte generadas artificialmente pasen el Test de Turing, en concreto AICAN –algoritmo abordado en el capítulo anterior. En concreto, el 75% de las personas encuestadas creyeron que se trataba de arte generado por un artista humano (Elgammal, 2018). Si la obra es capaz de transmitir emociones y significado a una audiencia, sería del todo reduccionista mantener una concepción del arte basada en una visión romántica del artista que expresa sus sentimientos a través de medios tradicionales.

En definitiva, no existe un consenso general sobre si las obras creadas mediante inteligencia artificial se pueden considerar arte. Sin embargo, bajo mi punto de vista existe un argumentario lo suficientemente relevante como para cuestionarnos la forma en la que se concibe actualmente el arte.

2. Autoría y problemas éticos en la aplicación del *Deep Learning* al arte

La aplicación de técnicas de *Deep Learning* –y de inteligencia artificial en general– al mundo del arte hace que nos enfrentemos a ciertas cuestiones éticas que resultan indispensables abordar.

2.1. Autoría y copyright de las obras generadas por inteligencia artificial

En primer lugar, nos encontramos con el debate de la autoría de las obras generadas mediante inteligencia artificial. En este contexto, no existe un consenso a nivel internacional, por lo que nos encontramos con una gran disparidad en cuanto las normativas nacionales.

En Estados Unidos, las obras generadas mediante inteligencia artificial no estarían protegidas por las normas de copyright. Desde la década de los setenta, la CONTU[15] mantiene que sólo pueden estar protegidas por copyright aquellas obras creadas por

[15] National Commission on New Technological Uses of Copyrighted Works

humanos, excluyendo las obras creadas por animales o por máquinas sin intervención humana (Samuelson, 2020).

En China, nos encontramos con el interesante caso *Feilin v. Baidu*. En este caso, el Tribunal manifestó que los *outputs* generados por inteligencia artificial no pueden ser protegidos mediante copyright. Sin embargo, sí podrá ser protegido por copyright si se lleva a cabo algún tipo de modificación del *ouput* por parte de un ser humano –en el caso de *Feilin*, colorear algunos dibujos (Samuelson, 2020).

En el caso de la Unión Europea, para que las obras de arte sean protegidas por copyright es necesario que cumplan con cuatro criterios: (i) relación con una producción en el campo literario, científico o artístico; (ii) ser el producto de un esfuerzo intelectual humano; (iii) ser el resultado de decisiones creativas que (iv) estén expresadas en el *output*. Si se cumplen estos requisitos, la autoría se atribuirá a la persona que haya contribuido desde un punto de vista creativo en el *output*, siendo en la mayoría de los casos el usuario que usa el modelo, no su programador (Hugenholtz & Quintais, 2021).

Sin embargo, también nos encontramos con países en los que los derechos de copyright se atribuyen al programador del modelo, notablemente en Reino Unido, Irlanda y Nueva Zelanda. En concreto, la sección 9(3) del Copyright, Designs and Patents Act de Reino Unido explica que en las obras generadas por ordenador, el autor será la persona que lleve a cabo los preparativos necesarios para la creación de la obra (Guadamuz, 2017).

Como se puede observar, no existe un consenso internacional acerca de la materia, por lo que nos podríamos encontrar con obras que en un país estarían protegidas por copyright, pero no podrían gozar de tal protección en otro.

2.2. Datasets y sesgos implícitos

La importancia de los datos que se usan como *input* en los modelos de *Deep Learning* que generan arte es primordial. Sin embargo, la selección de estos datasets no es –o no debería de ser– ajena a la ética.

En primer lugar, es necesario reflexionar sobre el papel de los artistas de las obras que se usan como *input* para este tipo de modelos. Si bien es cierto que existen implicaciones legales del uso de obras ajenas para entrenar modelos, existe un peligro real de abuso por

parte de compañías y plataformas que usan inteligencia artificial en el campo de las artes. Este tipo de compañías no suele dar demasiados detalles públicos sobre los datos que usan. En este sentido, existen sospechas de que Spotify estaría usando las canciones subidas a la plataforma para aplicar técnicas de inteligencia artificial sobre las mismas – sin consultar o reconocer a los artistas de las canciones *input* (Morreale, 2021). Las obras originales están protegidas por copyright, por lo que se necesitaría el permiso del artista para usar su obra como *input*. Sin embargo, en el contexto de la Unión europea existen dos excepciones: (i) uso para investigaciones científicas si tiene un acceso legítimo a los datos y (ii) uso más allá del científico salvo si el titular de los derechos restringe su uso de forma expresa (Sturm, Iglesias, Ben-Tal, Miron, & Gómez, 2019). Aunque su uso pueda ser legal, es indudable que los artistas y las grandes plataformas no se encuentran en pie de igualdad a la hora de negociar este tipo de cuestiones, pues al final del día el artista depende de las grandes plataformas para acceder a su público.

Al margen de los derechos de autor en los datasets usados, merece la pena detenerse en los posibles sesgos presentes en los modelos. Éstos aprenden de información pasada, por lo que tienden a perpetuar sesgos actuales. Desde el mundo académico se han llevado a cabo numerosos estudios que analizan la brecha de género en este campo. Por ejemplo, en el campo de los modelos de recomendación musical se ha observado que las plataformas de streaming tienden a penalizar a las artistas, dificultando su exposición a una mayor audiencia por culpa del bucle de *feedback* que se genera (Ferraro, Serra, & Bauer, 2021).

Además, los modelos de *Deep Learning* no son capaces de adelantarse a cambios culturales y tendencias. Si bien la aplicación de este tipo de técnicas puede parecer algo novedoso o futurista, al basarse en obras ya existentes existe peligro que de que se genere un arte de tipo conservador, que tienda poco a la innovación. Por ejemplo, el uso de técnicas de *Deep Learning* en la industria cinematográfica con herramientas como ScriptBook o Cinelytic puede llevar a penalizar proyectos novedosos y arriesgados, favoreciendo a aquellos que sigan fórmulas arquetípicas que se han llevado a cabo de forma recurrente. Esta lógica puede llegar a afectar a los propios profesionales de la industria, pues se favorecerá la contratación de actores, directores y guionistas de prestigio, restando oportunidades a personas que acaban de llegar a la industria (Raevskiy, 2020). Aquí también nos estaríamos enfrentando a un problema de sesgos. El 89% de las

nominaciones a los Oscar de la última década han caído en manos de personas blancas, y sólo el 28,8% de las nominaciones de las principales categorías pertenecen a mujeres (Sarkisian, 2021). Siendo este el estado actual de la industria en cuestiones de diversidad, si aplicamos técnicas de *Deep Learning* que se basan en experiencias pasadas para aprender, existe un claro peligro de perpetuar la tendencia y no cerrar las brechas de representación existentes.

2.3 Aspectos negativos para los artistas: La lucha por la atención del espectador

El uso generalizado de modelos capaces de crear obras artísticas puede tener repercusiones negativas para los artistas que actualmente forman parte de la industria. En capítulos anteriores ya se adelantó que el uso de este tipo de técnicas puede tener repercusiones en términos de empleo para la industria. Por ejemplo, si se reduce el tiempo necesario para editar un vídeo de 30 días a 24 horas gracias al *Deep Learning*, la demanda de trabajo para los profesionales caería de forma notable. Sin embargo, no sólo habría consecuencias para los perfiles técnicos de las distintas industrias del arte como editores de contenido o analistas de guiones en los estudios. Los artistas también se podrían ver directamente afectados, especialmente en lo que respecta al acceso a la audiencia.

En el ámbito de la industria musical, Spotify es una plataforma clave para los artistas, especialmente aquellos que no cuentan con muchos seguidores y quieren darse a conocer. Dentro de Spotify, es especialmente relevante la figura de las playlists elaboradas por la propia compañía, pues suponen una oportunidad de oro para aquellos artistas que buscan ampliar su audiencia. En 2019 saltó la noticia de que Spotify había llegado a un acuerdo con la productora sueca Epidemic Sound para crear música bajo perfiles de artistas falsos de cara a añadirla a sus playlists. Es decir, los compositores de la productora grababan canciones y las publicaban con nombres falsos de artistas inexistentes, a cambio de unos royalties menores que los que se pagarían a los artistas reales (Ingham, 2019). En este contexto, cabe asumir que los modelos de *Deep Learning* capaces de crear música agravarían esta tendencia.

Por supuesto, el fenómeno explicado no afectaría a los grandes artistas, pero podría tener serias implicaciones para los artistas pequeños. La popularización de los servicios de streaming supuso una democratización del acceso a canales de distribución en la

industria, abriendo las puertas a un gran número de nuevos artistas. Esto genera el conocido como *long tail problem,* en base al cual tenemos un reducido número de músicos populares y una gran mayoría que no tienen acceso a una gran audiencia. Así las cosas, nos enfrentamos a una situación en la que la música es más accesible al consumidor, pero los consumidores no van a reproducir más música por el hecho de tener más artistas disponibles. Por lo tanto, la probabilidad de que una canción sea reproducida por muchos consumidores disminuye. Si la generación de música mediante inteligencia artificial se generaliza, aumentaría más si cabe la competitividad en el mercado, agravando la situación de los pequeños artistas –que ya de por sí se encuentran en una situación difícil (Morreale, 2021).

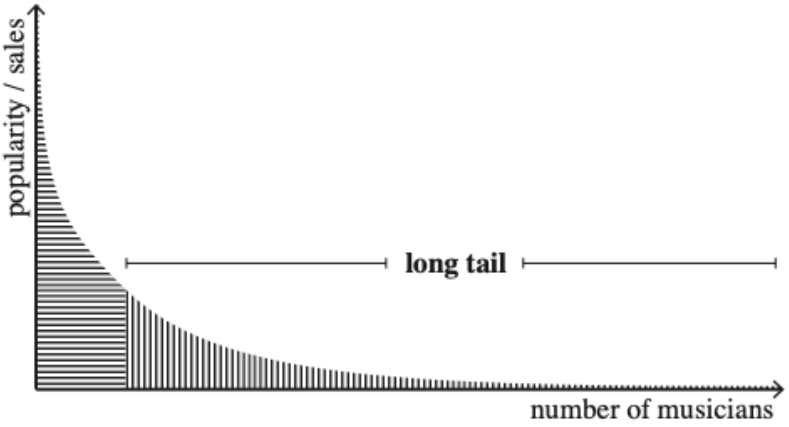


Figura 23
Long tail problem
Extraído de Morreale, F. (2021). Where Does the Buck Stop? Ethical and Political Issues with AI in Music Creation. Transactions of the International Society for Music Information Retrieval, 4(1), pp. 105–113

VI. Conclusión

En definitiva, el *Deep Learning* constituye uno de los campos más relevantes del *Machine Learning* en la actualidad. A lo largo del presente trabajo se ha podido comprobar cómo las características técnicas de las redes neuronales profundas hacen computacionalmente posible la creación de modelos para una infinidad de propósitos, contribuyendo al desarrollo de múltiples tecnologías, como son los vehículos de conducción automática y los asistentes digitales –por ejemplo, Amazon Alexa.

En concreto, se ha desarrollado una exposición de las principales aplicaciones del *Deep Learning* en el campo del arte. De esta forma, se ha observado cómo la aplicación de este tipo de técnicas tiene un verdadero potencial transformador en las distintas industrias del arte. Las nuevas herramientas que trae en *Deep Learning* permiten no sólo la creación de cuadros, canciones y poemas, sino que también permiten desarrollar una serie de instrumentos para actores de la industria que no ocupan un rol creativo, como son las aplicaciones que analizan guiones de cine y son capaces de pronosticar cómo sería recibido en el mercado o los modelos capaces de detectar canciones de cara a proteger los derechos de copyright.

Otra de las conclusiones claves a destacar es el hecho de que la aplicación del *Deep Learning* al arte no viene exenta de problemáticas a nivel ético y jurídico. Si bien el *Deep Learning* trae consigo oportunidades de negocio, el resultado no siempre será positivo para los artistas y profesionales del campo en términos de empleo y lucha por la atención del espectador. Además, presenta dificultades en términos de copyright para las obras creadas por estos medios.

En suma, el *Deep Learning* revolucionará el campo del arte, afectando a la práctica totalidad de los actores de sus diversas industrias. Sin duda, en los próximos años veremos surgir nuevos negocios que implementen soluciones novedosas en *Deep Learning* en este campo, además de mejoras en sus aplicaciones actuales. Sin embargo, es necesario no perder de vista las implicaciones éticas que esto acarrea.

Bibliografía

Abirami, S., & Chitra, P. (2020). The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases. *Advances in Computers vol. 117, Issue 1*.

Agarwal, S. (31 de Diciembre de 2017). *My thoughts on Skip-Thoughts*. Obtenido de Medium.com: https://medium.com/@sanyamagarwal/my-thoughts-on-skip-thoughts-a3e773605efa

Art, M. O. (2018). *Obvious Art.* Obtenido de http://obvious-art.com/wp-content/uploads/2020/04/MANIFESTO-V2.pdf

Artola Moreno, Á. (2019). Clasificación de imágenes usando redes neuronales convolucionales en Python.

Baheti, P. (29 de Noviembre de 2021). *Neural Style Transfer: Everything You Need to Know [Guide]*. Obtenido de V7: https://www.v7labs.com/blog/neural-style-transfer

Birtchnell, T. (2018). Listening without ears: Artificial intelligence in audio mastering. *Big Data & Society July–December 2018: 1–16*, 1.

Briot, J.P. (2021) From Artificial Neural Networks to Deep Learning for Music Generation - History, Concepts and Trends. Neural Computing and Applications, *Springer Verlag*, pp.39-65.

Brownlee, J. (2019). A Gentle Introduction to Pooling Layers for Convolutional Neural Networks. *Machine Learning Mastery https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/*.

Can Artificial Intelligence change the art market? (29 de Abril de 2019). Obtenido de Fintech News: https://www.fintechnews.org/can-artificial-intelligence-change-the-art-market/

Cetinic, E., & She, J. (2021). Understanding and creating art with AI: Review and Outlook.

Chauhan, N. K., & Singh, K. (2018). A Review on Conventional Machine Learning vs Deep Learning. *International Conference on Computing, Power and Communication Technologies (GUCON).* Greater Noida, UP, India.

Chollet, F. (2018). *Deep Learning with Python.* Shelter Island: Manning Publications.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. En *NIPS 2014 Workshop on Deep Learning, Diciembre 2014*

Cilimkovic, M. (2015). Neural networks and back propagation algorithm. *Institute of Technology Blanchardstown, Blanchardstown Road North Dublin*, 1-12.

Deng, Y., Tang, F., Dong, W., Ma, C., Huang, F., Deussen, O., & Xu, C. (2021). Exploring the Representativity of Art Paintings. *IEEE Transactions on Multimedia vol 23*, 2794-2805.

Desagulier, G. (25 de Abril de 2018). *Word embeddings: the (very) basics.* Obtenido de Around the word: A corpus linguist's research notebook: https://corpling.hypotheses.org/495

E.Kurt, D. (2018). Artistic Creativity in Artificial Intelligence. *Radboud University*.

Education, I. C. (14 de Septiembre de 2020). *Recurrent Neural Networks.* Obtenido de IBM Cloud Learn Hub: https://www.ibm.com/cloud/learn/recurrent-neural-networks

Elgammal, A. (19 de Octubre de 2018). *75% of people think this AI artist is human*. Obtenido de Fast Company : https://www.fastcompany.com/90253470/75-of-people-think-this-ai-artist-is-human

Ferraro, A., Serra, X., & Bauer, C. (2021). Break the Loop: Gender Imbalance in Music Recommenders. *CHIIR '21*, 249-254.

Find out fine art hottest trends. (s.f.). Obtenido de Kellify: https://kellify.com/tools/find-out-the-hottest-fine-art-trends/

Fu, Z., Tan, X., Peng, N., Zhao, D., & Yan, R. (2018). Style Transfer in Text: Exploration and Evaluation. *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 663-670.

Fukushima, K., & Miyake, S. (1980). Neocognitron: Self-organizing network capable of position-invariant recognition of patterns. *Proc. 5th Int. Conf. Pattern Recognition* (Vol. 1, pp. 459-461).

Gatys, L., Ecker, A., & Bethge, M. (2015). *A Neural Algorithm of Artistic Style.* Obtenido de arXiv:1508.06576

Ghazvininejad, M., Shi, X., Choi, Y., & Knight, K. (2016). Generating Topical Poetry. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, (págs. 1183-1191). Austin, Texas .

Guadamuz, A. (Octubre de 2017). *Artificial intelligence and copyright*. Obtenido de WIPO Magazine: https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html

Heathman, A. (2 de Septiembre de 2016). *IBM Watson creates the first AI-made film trailer – and it's incredibly creepy.* Obtenido de Wired: https://www.wired.co.uk/article/ibm-watson-ai-film-trailer

Huang, H., Jian, W., Wang, H., Zhu, X., Luo, W., Li, Z., . . . Liu, W. (2017). Real-Time Neural Style Transfer for Videos. *EEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7044-7052.

Hugenholtz, P. B., & Quintais, J. (2021). Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output? *IIC vol. 54 n.9*, 1191-1216.

IFPI. (2021). *Global Music Report.* Obtenido de https://www.ifpi.org/wp-content/uploads/2020/03/GMR2021_STATE_OF_THE_INDUSTRY.pdf

Ignjatić, J., Bajic, B., Rikalovic, A., & Culibrk, D. (2018). Deep Learning for Historical Cadastral Maps Digitization: Overview, Challenges and Potential. *26th International Conference on Computer Graphics, Visualization and Computer Vision*, (pág. 45). Plzen.

Ingham, T. (15 de Mayo de 2019). *'Fake Artists' Have Billions of Streams on Spotify. Is Sony Now Playing the Service at Its Own Game?* Obtenido de Rolling Stone: https://www.rollingstone.com/pro/features/fake-artists-have-billions-of-streams-on-spotify-is-sony-now-playing-the-service-at-its-own-game-834746/

Jhamtani, H., Gangal, V., Hovy, E., & Nyberg, E. (2017). Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models.

Kaleagasi, B. (9 de Marzo de 2017). *A New AI Can Write Music as Well as a Human Composer* . Obtenido de Futurism: https://futurism.com/a-new-ai-can-write-music-as-well-as-a-human-composer

Kiros, J. (14 de Noviembre de 2015). *neural-storyteller*. Obtenido de GitHub: https://github.com/ryankiros/neural-storyteller

Klausen, S. H. (2010). The Notion of Creativity Revisited: A Philosophical Perspective on Creativity Research. *Creativity Research Journal, 22(4)*, 347-360.

Lai, S., Liu, K., & He, S. (2016). How to Generate a Good Word Embedding. *IEEE Intelligent Systems, vol. 31. no.6*, 5-14.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature vol. 521*, 436.

Lee, C.-C., Lin, W.-Y., Shih, Y.-T., Kuo, P.-Y. P., & Su, L. (2020). Crossing You in Style: Cross-modal Style Transfer from Music to Visual Arts. *MM '20: Proceedings of the 28th ACM International Conference on Multimedia*, 3219-3227.

Los algoritmos ya venden cuadros por miles de euros. (1 de Enero de 2021). Obtenido de IMMUNE Technology Institute: https://immune.institute/los-algoritmos-ya-venden-cuadros-por-miles-de-euros/

Madhu, P., Kosti R., Mührenberg, L., Bell, P., Maier, A., Christlein, V. (2019). Recognizing character in art history using deep learning. *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritage Contents, p.15-22.*

Martínez, P. D. (2020). *Demostrador de aprendizaje reforzado en plataforma de juego.* Universidad de Cartagena - Trabajo de Fin de Grado.

Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones.* Rosario: Universidad Tecnológica Nacional – Facultad Regional Rosario Departamento de Ingeniería Química Grupo de Investigación Aplicada a la Ingeniería Química (GIAIQ) .

Mazzone, M., & Elgammal, A. (2019). Art, Creativity, and the Potential of Artificial Intelligence. *Arts, 8,* , 1-9.

McCaig, G, DiPaola, S., & Gabora, L. (2016). Deep convolutional networks as models of generalization and blending within visual creativity. *Proceedings of the 7th International Conference on Computational Creativity* (pp. 156-163). Palo Alto, CA: Association for the Advancement of Artificial Intelligence (AAAI) Press.

Mikalonytė, E. S., & Kneer, M. (2021). Can Artificial Intelligence Make Art? Folk Intuitions as to whether AI-driven Robots Can Be Viewed as Artists and Produce Art. 1-16.

Mordvintsev, A., Olah, C., & Tyka, M. (17 de Junio de 2015). *Inceptionism: Going Deeper into Neural Networks*. Obtenido de Google AI Blog: https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Morreale, F. (2021). Where Does the Buck Stop? Ethical and Political Issues with AI in Music Creation. *Transactions of the International Society for Music Information Retrieval 4(1)*, 105-112.

Music Informatics Group, G. T. (8 de Enero de 2018). *Neural Style Transfer for Musical Melodies*. Obtenido de Ashis Pati: https://ashispati.github.io/style-transfer/

Newitz, A. (25 de Abril de 2017). *An AI wrote all of David Hasselhoff's lines in this bizarre short film*. Obtenido de Ars Technica: https://arstechnica.com/gaming/2017/04/an-ai-wrote-all-of-david-hasselhoffs-lines-in-this-demented-short-film/

Newitz, A. (30 de Mayo de 2021). *Movie written by algorithm turns out to be hilarious and intense*. Obtenido de Ars Technica: https://arstechnica.com/gaming/2021/05/an-ai-wrote-this-movie-and-its-strangely-moving/

Odlen, I., Verma, P., Basica, C., & Kivelson, P. D. (2020). Transfer, Painting from Music using Neural Visual Style.

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

¿Qué es la masterización? (s.f.). Obtenido de LANDR: https://www.landr.com/es/que-es-la-masterizacion

Prateek, V., & Smith, J. (2018). *Neural Style Transfer for Audio Spectograms*. Obtenido de arXiv:1801.01589

Raevskiy, M. (7 de Octubre de 2020). *How AI Is Changing the Film Industry.* Obtenido de Medium.com: https://ai.plainenglish.io/how-ai-change-film-industry-a1514f450881

Robitzski, D. (18 de Julio de 2018). *Artificial Intelligence Is Automating Hollywood. Now, Art Can Thrive.* Obtenido de Futurism: https://futurism.com/artificial-intelligence-automating-hollywood-art

Samuelson, P. (2020). Legally Speaking AI Authorship? *Communications of the ACM vol. 63 n.7*, 20-22.

Sarkisian, J. (2 de Febrero de 2021). *A complete breakdown of the Oscars diversity problem over the past decade*. Obtenido de INSIDER: https://www.insider.com/oscars-diversity-problem-race-gender-numbers-academy-awards-2021-2

Schneider, T. (31 de Marzo de 2020). *A.I. Has the Potential to Change the Art Business—Forever. Here's How It Could Revolutionize the Way We Buy, Sell, and See Art*. Obtenido de Artnet News: https://news.artnet.com/market/ai-art-business-intelligence-report-2020-1812288

Sciforce. (8 de Junio de 2021). *3 Reasons to Use AI in Music Industry.* Obtenido de Medium.com: https://medium.com/sciforce/3-reasons-to-use-ai-in-music-industry-26486b3100bd

Seb. (2021). *An Introduction to Neural Network Loss Functions*. Obtenido de Programmathically: https://programmathically.com/an-introduction-to-neural-network-loss-functions/

Siegel, T. (8 de Enero de 2020). *Warner Bros. Signs Deal for AI-Driven Film Management System (Exclusive).* Obtenido de The Hollywood Reporter: https://www.hollywoodreporter.com/business/business-news/warner-bros-signs-deal-ai-driven-film-management-system-1268036/

Soleimany, A., & Amini, A. (19 de Febrero de 2021). *MIT Introduction to Deep Learning 6.S191: Convolutional Neural Networks.* Obtenido de Youtube: https://www.youtube.com/watch?v=AjtX1N_VT9E&ab_channel=AlexanderAmini

Soleimany, A., & Amini, A. (5 de Febrero de 2021). *MIT Introduction to Deep Learning 6.S191: Introduction to Deep Learning.* Obtenido de Youtube: https://www.youtube.com/watch?v=5tvmMX8r_OM&list=PLtBw6njQRU-rwp5__7C0oIVt26ZgjG9NI&ab_channel=AlexanderAmini

Soleimany, A., & Amini, A. (12 de 2021 de Febrero). *MIT Introduction to Deep Learning 6.S191: Recurrent Neural Networks.* Obtenido de Youtube: https://www.youtube.com/watch?v=qjrad0V0uJE&list=PLtBw6njQRU-rwp5__7C0oIVt26ZgjG9NI&index=2&ab_channel=AlexanderAmini

Sturm, B. L., Iglesias, M., Ben-Tal, O., Miron, M., & Gómez, E. (2019). Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis. *Arts, 8, 115*, 1-15.

Tamir, M. (26 de Junio de 2020). *What Is Machine Learning?* Obtenido de UC Berkeley: https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/

Torres, J. (22 de Septiembre de 2019). *Redes Neuronales Recurrentes.* Obtenido de Jordi TORRES.AI: https://torres.ai/redes-neuronales-recurrentes/

Vázquez, F. (21 de Diciembre de 2017). *Deep Learning made easy with Deep Cognition*. Obtenido de Medium.com: https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351

Vincent, J. (28 de Mayo de 2019). *Hollywood is quietly using AI to help decide which movies to make.* Obtenido de The Verge:

https://www.theverge.com/2019/5/28/18637135/hollywood-ai-film-decision-script-analysis-data-machine-learning

Vincent, J. (10 de Julio de 2019). *Runway ML puts AI tools in the hands of creators everywhere.* Obtenido de The Verge: https://www.theverge.com/2019/7/10/20682307/ai-machine-learning-easy-to-use-models-creatives-runway-ml

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deel Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 1-2.

Weng, S.-S., & Chen, H.-C. (2020). Exploring the Role of Deep Learning Technology in the Sustainable Development of the Music Production Industry. *Sustainability, 12, 625*, 2.

Zhang, J., Feng, Y., Wang, D., Wang, Y., Abel, A., Zhang, S., & Zhang, A. (2017). Flexible and Creative Chinese Poetry Generation Using Neural Memory.