

E-Personería y Vehículos Autónomos: un caso de estudio

Hugo Leonardo Rufiner^{1,2}

XIII Jornadas de Derecho Judicial:

“La Justicia ante los desafíos del mundo digital. Experiencias y perspectivas”.

Facultad de Derecho, Universidad Austral, Buenos Aires, septiembre 2020

¹ Instituto SINC-UNL-CONICET, ² Lab. Cibernética, FI-UNER

Resumen: *La Inteligencia Artificial ha abierto la posibilidad de crear máquinas autónomas, a las cuales es posible cederles el control de acciones que antes estaban reservadas exclusivamente a los humanos. Esta disciplina, actualmente en auge, consiste precisamente en el estudio y diseño de agentes o máquinas inteligentes. Estas máquinas poseen la capacidad para percibir su entorno, tomar decisiones y realizar acciones que maximizan sus posibilidades de éxito en tareas específicas. En este contexto se plantean los siguientes interrogantes que abordaremos en este trabajo: ¿Es posible pensar en las máquinas/agentes como “sujetos” en forma análoga a cómo lo hacemos con los humanos? ¿Podrían ejecutar entonces actos “voluntarios” similares a los nuestros? ¿Qué implicancias legales, éticas o filosóficas aparecen en estas situaciones? Finalmente se analiza brevemente como caso de estudio el de los vehículos autónomos actuales, las discusiones, la normativa y los posibles escenarios legales que se presentan en el caso de accidentes.*

Introducción

El tema de la interacción entre los hombres y las máquinas se podría decir que excede el mero ámbito de la acción humana porque aparecen elementos adicionales. No solo por la posibilidad de uso de máquinas por parte de las personas, sino también porque en el contexto actual estas máquinas pueden tener cierta autonomía y algún grado de interacción con los usuarios¹.

Actualmente la IA se ha difundido por todas partes y se ha incorporado en muchos dispositivos, facilitando las actividades diarias de las personas y produciendo una verdadera revolución en la sociedad. El hito más reciente ha sido la aparición del denominado *aprendizaje profundo* [12, 13, 15], que en conjunción con los avances en la capacidad del hardware y en la disponibilidad de grandes cantidades de datos, ha permitido resolver muchos problemas con un desempeño similar o incluso mejor que los humanos [10]. Esto se ha dado principalmente en el contexto de la IA no simbólica, en lo que se llama inteligencia artificial conexionista [4], que es la rama más biológicamente inspirada, dando lugar a lo que se conoce como redes neuronales artificiales profundas. Estas utilizan lo que se conoce acerca del cerebro y las neuronas biológicas para diseñar modelos neuronales interconectados con muchas capas que pueden aprender a resolver problemas.

Mediante este tipo de algoritmos ha sido posible dotar a las máquinas de ciertas capacidades que emulan diversos aspectos de la inteligencia humana, como es el caso que nos ocupa de la autonomía. Los vehículos autónomos, que pueden manejarse con poca o ninguna participación o asistencia de parte de un conductor humano están en prueba en distintos países desde hace varios años. Los primeros accidentes fatales ya han ocurrido, lo que plantea numerosas dudas sobre la conveniencia y seguridad de esta tecnología. Para tener una idea de la capacidad de estos vehículos lo mejor es verlos en funcionamiento en alguna demostración. Por ejemplo en el caso del auto de Google©, la persona sube al mismo y le indica en forma oral el lugar adónde desea ir, luego presiona un botón y el vehículo

¹ Este trabajo es un extracto del Capítulo: Rufiner, Hugo Leonardo, “Inteligencia artificial y acción humana” en Juan B. Etcheverry (ed.), “Perspectivas de la acción humana. Un abordaje interdisciplinario al problema de la comprensión del actuar humano”, Ed. Porrúa, México, enviado para publicación 2019.

realiza todas las operaciones correspondientes para llegar al sitio deseado sin ninguna otra intervención humana. Decide cuál es la ruta óptima, las acciones concretas y la velocidad en cada momento. El vehículo posee una representación interna del mundo y un conjunto de “reglas” para tomar estas decisiones a partir de los datos provistos por varios sensores y técnicas de procesamiento de imágenes².

Autonomía, agentes no humanos y E-Personería (Electrónica)

A partir de los desarrollos planteados se está discutiendo seriamente el tema de la posibilidad de otorgar algún tipo de personería electrónica para estos dispositivos [35]. Sería entonces posible utilizar el término de *agentes no humanos* para los dispositivos que evidencien diferentes grados de autonomía. El dispositivo podría ser bien una máquina física, vehículo, robot o también un agente virtual (algoritmo o sistema de software) que esté operando en internet o que ejecute algún tipo de acción (tanto física como virtual) con poca o ninguna supervisión humana.

Entonces podríamos preguntarnos si es lógico pensar en estos agentes como si fueran sujetos de forma análoga a nosotros. El concepto de agente que suele usarse en el contexto de la IA parece bastante cercano a la noción de sujeto que se tiene en las ciencias humanas y sociales. La pregunta sería entonces si estos agentes: ¿podrían ejecutar actos voluntarios como los realizados por los sujetos humanos?

Recordamos aquí, por cuestiones de completitud, las condiciones necesarias para que haya un acto humano voluntario en el contexto jurídico [31]. Las condiciones internas son discernimiento, intención y libertad, además que debe manifestarse por un hecho exterior. El discernimiento se define como la madurez intelectual para razonar, comprender y valorar el significado del acto y sus consecuencias. La intención está asociada al propósito de realizar el acto y se define como la correspondencia entre lo entendido y lo actuado. Finalmente la libertad consiste en la posibilidad de elegir entre varias opciones, sin coacción externa legítima. Partiendo de estas definiciones se puede replantear la pregunta en términos de la posibilidad o no de que se cumplan estas condiciones para el caso de las máquinas. O sea de si es posible o no hablar de “actos artificiales voluntarios”. Esta por supuesto sería pregunta abierta por lo que no se espera responderla completamente en este capítulo, sino más bien identificar algunos elementos que puedan ayudar a la discusión.

Se debe volver a remarcar aquí la importancia del uso, los conceptos y diferencias asociados con la terminología técnica específica de cada disciplina, como la filosofía, las ciencias jurídicas o la IA. En el caso específico de la IA el principal problema es que en muchos casos se ocultan una serie de simplificaciones muy importantes detrás de una jerga tecnológica profundamente influenciada por los antropomorfismos. Esto es propio de una disciplina que desde el principio se planteó como objetivo la emulación de la inteligencia humana, siendo esta última la referencia obligada con la cual compararse o incluso intentar equipararse. Cuando se habla por ejemplo de “agente inteligente” no es inteligente en el mismo sentido que los humanos somos inteligentes sino más bien en sentido analógico [40]. Aunque es cierto que algunos aspectos de nuestra inteligencia han sido satisfactoriamente imitados, a veces esta correspondencia pretende plantearse como “exacta”, al menos en forma implícita. Lo mismo ocurre cuando se habla de agente autónomo, no se trata de la misma autonomía con la cual nosotros actuamos. En la acción humana, según se entiende

² <https://www.youtube.com/watch?v=mP05-KPhahg>

actualmente, aparecen casi obligadamente cuestiones asociadas con una “perspectiva subjetiva”, de “primera persona”, “autopercepción” o “condiciones internas” como las ya discutidas.

Se revisará entonces el concepto de la autonomía para la IA, particularmente en el contexto de un reciente documento de la Unión Europea sobre este tema [35, 36]. En el mismo se utiliza como referencia la siguiente definición: un sistema es autónomo si es “capaz de operar en el entorno real sin ningún tipo de control externo, una vez que la máquina se activa y al menos en algunas áreas de operación, por largos períodos de tiempo”. En este contexto también se suele hablar de “grados de autonomía”. Por ejemplo, el caso del vehículo de Google mostrado anteriormente sería casi completamente autónomo o sin supervisión. Pero también existen sistemas de asistencia a la conducción con distintos grados de autonomía, como se verá más adelante.

Para la filosofía el concepto de autonomía estaría más relacionado con el “autogobierno”, o sea la capacidad de la persona de gobernarse a sí misma. Esto a su vez se relaciona con la independencia de la deliberación y elección de la persona sin la manipulación de otras personas. En este sentido el énfasis no está puesto tanto en cómo se lleva a cabo la tarea sino en el porqué. Especialmente en la volición, así como en la autenticidad de los objetivos que hacen que alguien actúe [36]. Del lado de la IA los fines son entonces más pragmáticos que del lado de la filosofía, como era de esperarse.

Si se aplicara esta noción de autonomía proveniente de la filosofía a un agente, robot o vehículo esto implicaría la posibilidad de que el mismo pudiera *establecer* y *decidir* sus propios objetivos. Otra vez la terminología utilizada en IA haría pensar, si se analiza sin mucho cuidado, que esto es perfectamente posible. Esta capacidad de decidir y establecer estaría también relacionada con la noción jurídica de libertad ya presentada. Sin embargo, desde mi punto de vista para que el agente posea una autonomía real es decir, que la máquina sea autónoma en el sentido en que los humanos lo son, se necesitaría de una mente consciente como la nuestra. Esto nos lleva a analizar la posibilidad de dotar de una mente consciente a estos agentes mediante la IA. Entiéndase “mente” en el sentido amplio de la palabra. Esta es una discusión en sí misma que ha sido abordada con matices por distintos autores, aquí se presentarán sólo algunos lineamientos generales [27, 30]. Para más detalles remitirse a nuestro trabajo reciente [40, 41].

Responsabilidad de las máquinas

Ahora que se ha mostrado que no existen actualmente máquinas conscientes, se retomará el tema de los vehículos autónomos, entendiendo esta autonomía en forma analógica y no equivalente a la humana. Más allá de la utilidad de este tipo de vehículo, su uso implica también algunos riesgos que deben ser cuidadosamente analizados. En este contexto surge inmediatamente la pregunta acerca de la responsabilidad sobre las acciones de estos vehículos, particularmente cuando producen algún tipo de daño.

Desde la aparición de los primeros VA se han presentado varios accidentes, una parte de ellos mortales. Por ejemplo, en uno de los últimos casos ocurrido en 2018 se cruzó una mujer peatón que llevaba una bicicleta en la noche y el vehículo la atropelló, lo que produjo su muerte (ver detalles en nota al pie n° 6, más adelante). ¿Cómo se trataría un caso como este?

Es posible pensar de acuerdo a lo que se ha presentado hasta ahora que la máquina resultaría desde el punto de vista jurídico “inimputable”. Es decir no se le puede imputar la responsabilidad del hecho porque carece de los elementos esenciales de una persona humana, como lo ya discutido acerca de la consciencia. Entonces ¿quién debe asumir la responsabilidad en esta situación? Esto no es sólo una cuestión legal sino que en realidad es también una cuestión técnica compleja. Porque se debe

determinar claramente la naturaleza del problema que produjo el accidente y aquí se mezclan factores humanos y tecnológicos. En general es necesario realizar una pericia técnica luego del incidente y por limitaciones de la tecnología actual resulta difícil y engorroso encontrar dónde estuvo el error y poder corregirlo.

Estas limitaciones tienen que ver con el método con el cuál se desarrollan los sistemas basados en la inteligencia artificial. Uno de los problemas es que intervienen muchas personas en su desarrollo, fabricación, puesta en funcionamiento, comercialización y uso, con diferentes roles y funciones a lo largo de todas las etapas. Por citar algunos: el diseñador del sistema, el programador, el científico de datos (que es una nueva área que surgió recientemente), el experto en seguridad, el que realiza las pruebas de calidad, el fabricante del hardware, el dueño que adquirió el vehículo, el que supervisa la tarea de conducción, etc. Y el otro problema es que el proceso de generación de estos algoritmos o agentes es distinto del enfoque clásico de desarrollo de sistemas informáticos que es más sencillo, lineal o en ciclos pero con la intervención constante de personas para tomar las decisiones de diseño. En el caso del diseño de los sistemas inteligentes hay una serie de características que los hacen bastante diferentes. En primer lugar el enfoque de diseño de los mismos suele ser casi totalmente guiado por datos, la intervención humana está en otros aspectos. Los algoritmos no están compuestos por reglas que propongan los desarrolladores o basado en el conocimiento de expertos, sino que el sistema extrae toda la información y las reglas mismas de los datos. En segundo lugar aparece el problema de la complejidad de los mismos sistemas, que en general son del tipo “caja negra” (por no poder conocerse con exactitud sus componentes, estructura o funcionamiento interno). Están compuestos por la interconexión de muchos elementos, lo que produce funciones anidadas complejas no lineales y con parámetros aleatorios, por lo que no resulta sencillo predecir el efecto de modificar algún parámetro. Se dice que no son interpretables, en el sentido que su funcionamiento particular no puede ser explicado en forma directamente comprensible por los humanos. Por ejemplo, en el caso del vehículo que chocó al peatón ya discutido es muy difícil saber cuál de los cientos de miles de parámetros o elementos que hay en el sistema fue el “responsable” de la falla. De hecho seguramente no fue uno solo sino que suele ser un problema distribuido entre varios parámetros o elementos del sistema. La otra característica es que suelen ser adaptativos en relación con los cambios de los datos que reciben del ambiente y pueden modificar sus parámetros para aprender cómo actuar frente a estos cambios. Por ejemplo si alguien compra uno de estos vehículos viene con una configuración standard de fábrica, pero se puede adaptar a las condiciones del entorno cercano a la vivienda del dueño y sus rutas cotidianas. Además, el fabricante puede mandar actualizaciones periódicas del software por cuestiones de mejoras de funcionalidad o de seguridad ya que puede estar sujeto a ataques, de virus informáticos o hackers. Todo esto causa una serie de complejidades al abordaje de un problema, tanto del punto de vista técnico como jurídico.

Inteligencia artificial interpretable

A raíz de los inconvenientes mencionados ha surgido el área de la *inteligencia artificial interpretable* que tiene por objetivo por diseñar nuevos sistemas inteligentes que puedan explicar o dar cuenta del porqué de sus acciones. Los primeros sistemas inteligentes basados en reglas, tales como los sistemas expertos, eran más fácilmente explicables. Por ejemplo, en un problema de clasificación de frutas se podía decir que si la variable x asociada al tamaño, era mayor que un valor dado entonces se trataba de una manzana y si era menor entonces se trataba de una frutilla. Esto es porque las reglas eran interpretables de manera directa, pero a medida que se van tratando de resolver problemas más difíciles, se necesitan más datos y modelos más complicados, por lo que se va perdiendo esta explicabilidad. Ya no se tienen reglas sino un cúmulo de elementos todos interconectados. El problema no es sólo averiguar la causa de una falla sino también cómo explicarla a un tercero que no sea experto en la tecnología empleada. Por ejemplo, en el caso de una pericia que se quiere explicar

el resultado de una determinada prueba al juez, o en una compañía de software donde un gerente necesita entender cómo funciona un sistema por motivos de confianza y responsabilidad.

En cualquier caso para analizar un accidente no basta con que los métodos sean interpretables, sino que se necesita también recolectar muchos datos por lo que aparece la necesidad de grabar diferentes variables a lo largo del funcionamiento de los dispositivos. Esto es similar a la idea de las *cajas negras* de los aviones³, donde se graban constantemente todas las variables importantes de vuelo del aparato, ambientales y las conversaciones en la cabina. Después, frente a un accidente, se puede reconstruir la situación para tratar de averiguar qué lo ocasionó. De hecho, algunas regulaciones ya piden en forma obligatoria este equipamiento extra para los VA y obviamente esto encarece el costo de los dispositivos pero mejora la seguridad. La IA interpretable es un tema de investigación actual que está volviendo a incorporar enfoques basados en lógica simbólica y métodos estadísticos, para buscar la mejor forma de extraer reglas a partir, por ejemplo, del conocimiento distribuido almacenado en las conexiones y pesos de una red neuronal artificial. Otra forma es mostrar qué características de los datos son las que el algoritmo está considerando como más importantes para tomar una decisión determinada. Por ejemplo, en el contexto de una tarea de clasificación de animales, se pueden resaltar las zonas de una foto que el algoritmo asocia como más importantes para decidirse por una especie en particular y a partir de allí es más sencillo detectar algún problema.

Regulaciones y estándares existentes para los vehículos autónomos

En cuanto a las regulaciones hay un gran vacío, en la mayoría de los países no hay nada todavía. EEUU es uno de los países que más legislación tiene sobre el tema, y los que siguen son Japón, Corea del Sur y algunos países de la Unión Europea que vienen discutiendo este tema desde hace varios años [32, 35, 36, 37, 38, 39]. Existen varios estudios comparados acerca de estas regulaciones (como por ejemplo [37]). Podemos decir rápidamente que aparecen distintas variantes, en algunos casos se permite la circulación sólo con fines de prueba y en condiciones limitadas, en otros sólo si se cuenta con sistemas especiales de seguridad (por ejemplo disponer de un botón de emergencia para volver a la función manual), también se pide a veces que la persona que actúa como guardián o como asistente a la conducción tenga una licencia especial o algún tipo de seguro especial que se contrata sobre todo por los posibles daños a terceros.

En la cuestión de la autonomía y la responsabilidad también hay distintos enfoques, algunos pasan a tratar al dueño con una figura similar a la del *tutor* de la máquina o en otros casos tratándolo como al que posee un animal o una herramienta que puede resultar peligrosa. Estas son algunas de las figuras que se toman a partir de analogías con legislaciones de otros campos para tratar de aplicarlas a este dominio de los VA.

El tema de la personería electrónica ha sido discutido extensamente en la UE, y básicamente existen dos enfoques aplicables al caso de los vehículos autónomos [35, 36]. O se los trata como mercancías (i.e. por ejemplo equivalente a un lavarropas inteligente sofisticado) o bien como un agentes no humanos, donde estaríamos más cerca de las discusiones que se han planteado aquí. Un dato interesante es que la respuesta de la comisión que estudio este tema específicamente ha sido que: “La tecnología aún no está lista para ofrecer el grado de autonomía como para otorgar personalidad o estatus legal a los robots. Sin embargo, los problemas éticos, legales y sociales relacionados con los sistemas futuros, dotados de más inteligencia y autonomía, son tomados seriamente en cuenta por la comunidad de sistemas robóticos y sistemas cognitivos artificiales” [35].

³ Estas no se denominan cajas negras por desconocerse su contenido, como en el uso anterior en este mismo capítulo, sino por el color con el que se pintaban originalmente.

Otro aspecto importante son los niveles de autonomía, aunque el término técnico usado en los estándares es el de automatización (derivado de su uso en los pilotos automáticos). Como ya se indicó hay vehículos que no poseen ninguna automatización hasta otros donde no interviene de ninguna forma el humano y entre medio existen varios niveles. Existen dos tipos de estándares el de la SAE-I (Sociedad Internacional de Ingenieros Automotrices⁴) y el de la NHTSA (Administración Nacional de Seguridad Vial de EEUU⁵), que son importantes para determinar la cuestión de la responsabilidad civil. Para tener como referencia a continuación se detallan los niveles de automatización según cada caso. Para la SAE-I son 6 niveles, en los primeros tres niveles interviene el conductor humano: 0 (sin automatización), 1 (asistencia a la conducción) y 2 (automatización parcial); mientras que en los tres niveles más altos el vehículo posee un sistema de conducción automatizado más sofisticado: 3 (automatización condicional), 4 (automatización alta) y 5 (automatización completa). Mientras que para la NHTSA son 5 niveles: 0 (sin automatización), 1 (automatización de funciones específicas), 2 (automatización de funciones combinadas), 3 (automatización de auto-conducción limitada) y 4 (automatización de auto-conducción completa).

Otra cuestión a tener en cuenta es la de establecer alguna referencia cuantitativa para medir la seguridad de la conducción. En el caso de vehículos bajo control humano tradicional esto se ha medido y es del orden de 1,18 muertes por cada 160.000.000 de kilómetros conducidos. Sin embargo, dado que todavía no se cuenta con suficientes estadísticas de conducción para el caso de los vehículos autónomos y además que estos se han ido perfeccionando, no es posible responder actualmente a la pregunta de si resultan más o menos seguros que los conducidos por humanos. A las fecha se han registrado sólo 5 accidentes fatales⁶. Aún se requieren muchas más pruebas, experimentación, recolección de datos y ajustes antes de poder proporcionar de manera convincente niveles de seguridad claros y demostrablemente más altos que los de los vehículos tradicionales.

Para terminar se analizarán brevemente las características del caso del primer atropello mortal protagonizado por un vehículo autónomo⁷. El accidente ocurrió la noche del domingo 18 de marzo de 2018, una mujer de 49 años murió en Tempe, Arizona (EEUU) al ser atropellada por un vehículo operado por UBER, según informó a través de un comunicado la policía local: "El vehículo se dirigía hacia el norte cuando una mujer que caminaba fuera del paso de peatones cruzó la calzada de oeste a este y fue arrollada por el vehículo de Uber". El vehículo que protagonizó el accidente estaba en el modo autónomo, aunque había una persona en su interior tras el volante como supervisor o auxiliar. Se trataba de un auto Volvo XC90 con una serie de modificaciones que realizó la compañía para convertirlo en autónomo. En el exterior posee una caja con sensores y cámaras sobre el techo y en su interior varias pantallas para tener mejor control de la situación. Este caso fue estudiado por casi un año y se determinó que el conductor se había distraído. El vehículo no tenía automatización completa, sería equivalente a SAE-I nivel 4, porque necesitaba que el conductor este siempre disponible para tomar el volante en el caso de que el sistema no actúe. El fabricante lo declara expresamente en la documentación del vehículo. El informe de la pericia consignó que un programa del auto detectó a la mujer "unos seis segundos antes del impacto", cuando el vehículo viajaba a 69 kilómetros por hora, y que el software "clasificó a la peatona como un objeto desconocido, como un vehículo, y luego como una bicicleta con diferentes expectativas sobre la ruta a seguir. También añade que la mujer vestía ropas oscuras y empujaba una bicicleta que no tenía reflectores laterales cuando cruzó una sección de la carretera que no estaba iluminada. El informe dice que el sistema de conducción

⁴ <https://www.sae.org>

⁵ <https://www.nhtsa.gov>

⁶ https://en.wikipedia.org/wiki/List_of_self-driving_car_fatalities

⁷ https://elpais.com/tecnologia/2018/03/19/actualidad/1521479089_032894.html

autónoma determinó que se necesitaba un frenado de emergencia 1,3 segundos antes del impacto, pero que para entonces ya era demasiado tarde. La persona que iba en el asiento del conductor frenó menos de un segundo antes del impacto, con lo que fue incapaz de evitar a la mujer peatón. Por estas razones finalmente se le atribuyó la responsabilidad civil al conductor. Aunque Uber no fue declarado culpable en el accidente, la compañía si se vio obligada a terminar con sus pruebas de autos sin chofer en Arizona y despidió a 300 conductores.

Conclusiones

El grado de desarrollo de la tecnología actual ha permitido crear máquinas inteligentes a las cuales es posible cederles el control de acciones que antes estaban reservadas a los humanos. La inteligencia artificial consiste precisamente en el estudio y diseño de agentes inteligentes con capacidad para percibir su entorno, decidir y actuar en consecuencia. Los recientes avances logrados en esta área se han debido a una serie de ingeniosos trucos que han permitido una mejora notable en los algoritmos y en los métodos de optimización. A partir de estas técnicas y las grandes cantidades de datos disponibles es posible descubrir, modelar y analizar relaciones cada vez más complejas. Un ejemplo de esto es el caso de los algoritmos que controlan los vehículos autónomos, que desde hace un tiempo han comenzado a utilizarse en distintos lugares del mundo. En este trabajo se han analizado las implicancias técnicas, filosóficas y legales de la utilización de este tipo de dispositivos.

Bibliografía

- [1] W. S. McCulloch, W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, Vol. 5, Issue 4, pp. 115-133 (1943).
- [2] Turing, Alan, "Computing Machinery and Intelligence", *Mind*, Vol. LIX, N° 236, pp. 433–460, (1950).
- [3] Searle, John, "Minds, Brains and Programs", *Behavioral and Brain Sciences*, Vol. 3, N° 3, pp. 417–457 (1980).
- [4] Christoph von der Malsburg, "Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms" (1986).
- [5] Lecun, Y., "Modeles connexionnistes de l'apprentissage (connectionist learning models)". PhD thesis. Université P. et M. Curie, Paris 6 (1987).
- [6] Cybenko, G., "Approximations by superpositions of sigmoidal functions", *Mathematics of Control, Signals, and Systems*, 2(4), 303-314 (1989).
- [7] Kurt Hornik, "Approximation Capabilities of Multilayer Feedforward Networks", *Neural Networks*, 4(2), 251–257 (1991).
- [8] Mikel Olazaran, "A Sociological Study of the Official History of the Perceptrons Controversy", *Social Studies of Science*, Vol. 26, No. 3, pp. 611-659 (1996).
- [9] Minsky, Marvin, "The Emotion Machine: From Pain to Suffering", *Proc. of the ACM Conference on Creativity and Cognition*, ACM Press (1999).
- [10] Yoshua Bengio, "Learning Deep Architectures for AI", *Foundations and Trends in Machine Learning archive*, Vol. 2 Issue 1, pp 1-127 (2009).
- [11] Russell, S. and Norvig, P., "Artificial Intelligence: A modern approach". Third Edition, Prentice Hall (2010).
- [12] Edwin Chen, "Introduction to Restricted Boltzmann Machines" (2011).
- [13] Marc' Aurelio Ranzato, "Neural Nets for Vision", *Tutorial on Deep Learning* (2012).
- [14] Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test, *Time*, June 9, (2014), <http://time.com/2847900/eugene-goostman-turing-test/>
- [15] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, "Deep learning", *Nature* vol. 521, pp. 436–444 (2015).
- [16] V. Mnih et al., "Human-level control through deep reinforcement learning", *Nature* vol. 518, pp. 529–533 (2015).
- [17] Leon A. Gatys, Alexander S. Ecker and Matthias Bethge, "A Neural Algorithm of Artistic Style" (2015).
- [18] A. Elgammal and B. Saleh, "Quantifying Creativity in Art Networks", *6th Int. Conf. on Computational Creativity (ICCC)*, USA (2015).

- [19] Ian Goodfellow, Yoshua Bengio and Aaron Courville, “Deep Learning”, MIT Press, (2016) <http://www.deeplearningbook.org>
- [20] Google’s ‘Inceptionism’ Art Sells Big at San Francisco Auction, Artnet News, (2016), <https://news.artnet.com/market/google-inceptionism-art-sells-big-439352>
- [21] Aäron van den Oord et al, “Wavenet: a generative model for raw Audio”, Google DeepMind, London, UK (2016).
- [22] Harry Potter: Written by Artificial Intelligence, Max Deutsch, (2016), <https://medium.com/deep-writing/harry-potter-written-by-artificial-intelligence-8a9431803da6>
- [23] Twitter taught Microsoft’s AI chatbot to be a racist asshole in less than a day, The Verge (2016), <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>
- [24] D. Silver et al, “Mastering the game of Go without human knowledge”, Nature vol. 550, pp 354–359 (2017).
- [25] Zhang et. al., “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks”, ICCV (2017).
- [26] I. Gareis, L. Vignolo, R. Spies and H. L. Rufiner, “Coherent Averaging Estimation Autoencoders applied to evoked potential processing”, Neurocomputing (2017).
- [27] Tononi, Boly, Massimini and Koch, "Integrated information theory: from consciousness to its physical substrate". Nature Reviews Neuroscience. 17 (7): 450–461 (2017).
- [28] The Ethics and Governance of Artificial Intelligence, Course MIT, Media Lab (2018), <https://www.media.mit.edu/courses/the-ethics-and-governance-of-artificial-intelligence/>
- [29] Blue Brain Project, EPFL, Swiss (2018), <https://bluebrain.epfl.ch/>
- [30] Is Consciousness Entirely Physical? Interview Series, (2018), <https://www.closetotruth.com/series/consciousness-entirely-physical>
- [31] Código Civil y Comercial de la Nación Argentina, Título IV, “Hechos y actos jurídicos”, Capítulo 1, Art. 260 (2014).
- [32] Varela, Agustín, “Los robots autónomos y el problema de la atribución de resultados lesivos”. Revista de derecho penal y procesal penal. Buenos Aires (2018).
- [33] Chuanyi Li et. al. “A Novel Convolutional Neural Network for Statutes Recommendation”, Proc, 15th Pacific Rim International Conference on AI, China, August 2018.
- [34] “Comparing the Performance of AI to Human Lawyers in the Review of Standard Business Contracts, LawGeex Reports, Feb. 2018.
- [35] Parlamento EUROPEO: Asunto: Derechos de los robots. Documento E-011289-13 (2013).
- [36] “Suggestion for a green paper on legal issues in robotics”, euRobotics The European Robotics Coordination Action, Grant Agreement N° 248552, Dec. 31, 2012.
- [37] Danesi, Cecilia C., “Inteligencia artificial y responsabilidad civil: un enfoque en materia de vehículos autónomos”, La Ley – Suplemento Especial, Noviembre (2018) .
- [38] Fernández, Diego, “El impacto de la inteligencia artificial en el derecho”, La Ley, Octubre (2017) .
- [39] Muñiz, Carlos, “Para nosotros, para nuestra posteridad, y para todos los robots del mundo que quieran habitar el suelo argentino: ¿puede la inteligencia artificial ser sujeto de derecho?”, Revista Código Civil y Comercial, N° IV, Julio (2018).
- [40] Soler Gil, Francisco José, “¿Puede la inteligencia artificial dar una “mente” a las máquinas? Reflexiones y preguntas desde la filosofía”, Semana de Investigación Interdisciplinar “Del yo a la persona”, Universidad Austral, Buenos Aires, Agosto de 2018.
- [41] Rufiner, Hugo Leonardo, “¿Puede la inteligencia artificial dar una “mente” a las máquinas?” (Algunas respuestas desde la ciencias de la computación), Semana de Investigación Interdisciplinar “Del yo a la persona”, Universidad Austral, Buenos Aires, Agosto de 2018.
- [42] Reyes A; P. Rosso, D. Buscaldi, “From humor recognition to irony detection: The figurative language of social media”, Data & Knowledge Engineering 74 (2012) 1–12.
- [43] Etzioni, Oren, “When will Superintelligence arrive?”, Instituto Allen de Inteligencia Artificial, Universidad de Washington (2016).