Riesgos del desarrollo y utilización de la inteligencia artificial

Introducción

Los avances tecnológicos de la última década han liderado a la humanidad a una época sin precedentes en la historia. Dentro de estos últimos avances se encuentran los realizados en el campo del software y la electrónica, campos que actualmente controlan y están presentes en la totalidad de la vida cotidiana, ya que están presentes en la gran mayoría de aparatos electrónicos, lo que incluye el manejo de sistemas productivos, bancos o la bolsa de comercio.

Ahora, como es mencionado por [Gonçalo](http://todosibuc.uc.cl/primo_library/libweb/action/search.do?vl(freeText0)=Carri%c3%a7o%2c+Gon%c3%a7alo&vl(1080283163UI0)=creator&vl(1080299580UI1)=all_items&fn=search&tab=libros_tab&mode=Basic&vid=56PUC_INST&scp.scps=scope%3a(56PUC_INST)%2cscope%3a(puc_alma)%2cscope%3a(puc_dspace)%2cprimo_central_multiple_fe&ct=lateralLinking) Carriço, “la inteligencia artificial comprende la respuesta a la pregunta de Alan Turing en 1950: “¿Can machines think?”, […] más recientemente se ha definido como la máquina que puede resolver problemas complejos resueltos por un humano con su inteligencia natural” (Carriço, 2018), es por esto que el hombre, buscando ayuda para resolver problemas cada vez más difíciles se ha visto impulsado a crear este tipo de tecnologías, en la actualidad se han llevado a cabo muchos avances en este campo, sin embargo, como se menciona en el World Policy Journal,“los avances en IA pueden traer oportunidades para gente y negocios sin precedentes, pero estos avances vienen con riesgos hasta ahora largamente ignorados […], negando la preocupación fundamental: mantener el control sobre sistemas que algún día sean más inteligente que los humanos.” (Winter-Levy y Trefethen, 2016). Es precisamente la despreocupación por el control de esta tecnología lo que trae consigo todos los riegos de la Inteligencia Artificial; ¿cómo se puede mantener control a pesar de sea más inteligente que nosotros?, ¿qué riesgos trae consigo el que sea liberada al mundo real?

Para responder las interrogantes se pretende describir los riesgos en el desarrollo y la utilización de la inteligencia artificial, y discutir sobre algunas soluciones a estos problemas presentadas dentro de los últimos 5 años. Para ello, primero se presentan las características, etapas y riesgos del desarrollo de la inteligencia artificial, luego se describen algunas aplicaciones tecnológicas, junto a los riesgos que estas traen. Además, se pretende recabar las estrategias que mitiguen estos riegos, encontrando de estas, las más urgentes e importantes de implementar.

Desarrollo

El cerebro y la máquina

Antes de comenzar, se es mencionada la etapa tecnológica precursora a la inteligencia artificial, que corresponde al periodo comprendido entre la invención de la computadora, y el desarrollo de las primeras formas de I.A. Se encuentra dentro de este periodo que las diferencias entre el funcionamiento del cerebro humano y la computadora son presentadas por Whitworth: “Las computadoras actuales […] tienden a centralizar el control más que distribuirlo, usar un procesamiento secuencial más rápido que procesamiento en varios procesos en paralelo […]” (Whitworth, 2009), por lo que es posible afirmar que las principales diferencias corresponden al método básico con que se procesa la información y la cantidad de procesos simultáneos posibles.

Gracias a lo anterior es posible deducir que para desarrollar la inteligencia artificial, no solo basta el desarrollo de software complejos, sino que tambien desarrollo en la tecnología de procesamiento de la información, y de electrónica computacional, ya que estos corresponden a las bases del funcionamiento computacional en general, tal cual es afirmado por Hwang: “[…] es un proceso computacional, y es por esto que está enlazado con el poder computacional del material tangible de los chips y semiconductores en los cuales se está operando.” (Hwang, T., 2018).

Máquinas que aprenden solas

La primera etapa del desarrollo de la inteligencia artificial corresponde a la etapa donde se encuentra desarrollada una inteligencia artificial temprana y de corto alcance, también denominada inteligencia artificial débil, impulsada por el proceso computacional denominado “machine learning”, que es identificado por Tim Hwang como “el estudio de algoritmos que se mejoran a sí mismos a través de información” (Hwang, T., 2018), estudio cuya aplicación a sistemas computacionales permite que gracias a la velocidad de procesamiento de información y el enfoque en procesos individuales se obtengan resultados claros en un periodo finito de tiempo. Huang Gao (2014) afirma que la técnica es implementada en diversos campos, como ingeniería biomédica o sistemas de identificación, control o robóticos, sin embargo, corresponde a una temprana aproximación a la real capacidad de la inteligencia artificial, ya que es capaz solo de desempeñar una tarea específica, y para su desarrollo requiere una gran cantidad de información previa de la acción objetivo.

Una aplicación real este nivel corresponde a AlphaGo, inteligencia artificial desarrollada por la empresa DeepMind. Dentro de lo publicado por el equipo desarrollador en su artículo (Silver, D., Huang, A., …, Graepel, T. y Hassabis, D., 2016), se tenía por objetivo crear un programa computacional capaz de jugar Go, un juego clásico asiático, a partir de “Machine learning” con base de juegos y estrategias existentes previamente, llegando así a lograr vencer al campeón europeo el 100% de las veces jugadas, siendo así el primer programa computacional en lograrlo. Este ejemplo muestra, por un lado, el alcance posible de la inteligencia artificia, y por otro, la limitación de sólo aplicar esta técnica con una acción especifica como objetivo.

El principal riesgo de la inteligencia artificial en esta etapa corresponde a los virus computacionales, los cuales, como se afirma en el paper de Alexey Turchin y David Denkenberger:

” Se ve probable que, en el futuro los virus sean más sofisticados que los contemporáneos, y tendrán implementada una inteligencia artificial. Esto puede incluir la capacidad de modelar el mundo exterior y adaptar su comportamiento a este. Los virus que implementen inteligencia artificial pueden llegar incluso a imitar el lenguaje humano […]” (Turchin y Denkenberger, 2018).

Indicando esto que, a pesar de ser la etapa inicial de la inteligencia artificial, se presenta un gran riesgo para los sistemas financieros y gubernamentales basados en sistemas informáticos.

Además de lo anterior, la implementación de tecnología que utilice machine learning, como por ejemplo simulaciones o software de producción adaptable, trae consigo consecuencias económicas, ya que de implementar estas, se requiere de inversiones considerables en computadoras del mejor rendimiento, generando así mayor consumo por producto electrónicos de alta calidad, sin embargo, tambien genera desplazamiento laboral producto de la automatización. Esto es concluido por Hwang: “Avances en simulaciones, juegos y programas adaptables, son factores probables de impactar y expandir el mercado, pero tambien el acelerar el desplazamiento de los trabajos.” (Hwang, T, 2018).

Inteligencia artificial al nivel humano

La segunda etapa del desarrollo de la inteligencia artificial corresponde a una inteligencia artificial joven o tambien denominada inteligencia artificial de alto nivel, que es definida como “I.A. de un nivel por sobre la mayoría de las personas, pero por debajo del nivel de una súper inteligencia” (Turchin y Denkenberger, 2018). Se estima en el estudio realizado por Katja Grace (Grace, Salvatier, Dafoe, Zhang y Evans, 2017) que existe un 50% de probabilidad de que para el año 2060 se alcance este nivel de desarrollo de la inteligencia artificial, la que sería capaz de sobrepasar la capacidad humana totalmente. Para un análisis más detallado, la inteligencia artificial de esta etapa se ha de dividir según la capacidad de auto mejorar reescribiendo su código base.

La primera I.A. de este nivel que presenta menor riesgo es la que no es capaz de auto mejorar, ya que como firman Turchin y Denkenberger (2018), “Muchos de los riesgos de la inteligencia artificial de nivel humano son similares a los riesgos de una inteligencia artificial débil.”, siendo estos los ya descritos anteriormente, sin embargo, de masificarse estas eclipsando a los humanos reales, se encuentra otro riesgo, este es tambien descrito por Turchin y Denkenberger (2018): “Un riesgo específico es que las “Máquinas humanas” serán filosóficamente zombis. En este caso, de tener un mundo donde la mayoría de los seres sean de esta naturaleza, el mundo aparecerá ser feliz y disfrutable […] Pero carecería de experiencia subjetiva, estando de hecho, subjetivamente muerto.” Para ejemplificar este tipo de IA, se tiene lo presentado por Carrico (2018): “Un ejemplo de este tipo de IA puede ser visto en films de ciencia ficción, en los robots como C-3PO o R2-D2 de la saga Star Wars”, teniendo así que este tipo de inteligencia artificial está presente en el arte y la cultura.

La segunda I.A. de este nivel es la que presenta la mayor cantidad de riesgo, ya que, al tener la capacidad de modificar su código, posee la capacidad de liberarse de restricciones y medidas de control, alcanzando de esta forma lo que Turchin y Denkenberger (2018) describen: “En un comienzo rápido, una I.A. obtiene la dominación mundial en semanas o meses; en un comienzo lento, muchas I.As. evolucionan simultáneamente”, ejemplificando de esta forma el posible alcance que esta IA puede llegar a lograr.

Dentro de los principales riesgos de la segunda inteligencia artificial que se tiene un comienzo lento y, por lo tanto, un aumento paulatino en sus capacidades, lo que es descrito por Shabbir y Anwer (2015): “Los riesgos mayores de la inteligencia artificial incluyen una carrera armamentística autónoma, […], desempleo, discriminación social y desigualdad de poder en las sociedades. […] Dado el debido tiempo, la inteligencia artificial realizará mejor las tareas en comparación a los humanos, es por esto la pérdida de empleos.”, de esto se puede afirmar que las consecuencias pueden llegar a ser catastróficas para la sociedad, amenazando así el sistema económico base de una sociedad.

Dadas las habilidades de una inteligencia artificial en esta etapa, las aplicaciones de comprenden la totalidad de las actividades humanas, tal cual afirma Shabbir y Anwer (2015): “En el lado positivo, el avance tecnológico producido por la inteligencia artificial ayudará a erradicar enfermedades y prevenir guerras […] En el corto plazo reemplaza gente y cambia la naturaleza del trabajo que los humanos realizan. En el largo plazo, crea nuevos tipos de trabajo.”, por lo anterior, se deduce que el mayor beneficio de la inteligencia artificial está en el avance tecnológico e industrial.

La super inteligencia

La última etapa que la inteligencia artificial puede alcanzar corresponde a una inteligencia artificial madura o super inteligencia artificial, que es definida por Turchin y Denkenberger (2018) como la consecuencia de permitir a una inteligencia artificial mejorarse a sí misma el tiempo suficiente para alcanzar un nivel de inteligencia capaz de controlar toda la tierra. Un ejemplo de esta inteligencia artificial correspondería a una inteligencia artificial dominante del tipo singleton, que es definida por Bostrom (2005) como un ente capaz de tomar de decisiones de alto nivel, prevenir amenazas y ejercer control absoluto sobre su territorio de dominio.

Los riesgos de la inteligencia artificial a este nivel se pueden diferenciar a partir del cómo esta inteligencia artificial esté presente en el mundo, todos estos son descritos por Turchin y Denkenberger (2018),

La mayor aplicación de esta I.A. corresponde a

Los métodos de seguridad

Dentro de los métodos de seguridad, se tiene que para cada etapa se requieren nuevos y distintos mecanismos, estos de desarrollan a continuación.

La seguridad en la primera etapa del desarrollo, según Mathew Scherer (2016), para evitar el uso indiscriminado de la I.A. o con malas intenciones, propone regular legalmente el desarrollo y venta de estos productos, para que de esta manera el precio del mercado refleje los riesgos asociados, y además se pueda castigar duramente a los desarrolladores que no realicen su trabajo bajo las normas de seguridad básicas, desincentivando así esta práctica. De la misma manera, Shabbir y Anwer (2015) proponen “desarrollar un sistema efectivo de coordinación y cooperación para construir un ambiente competente en que el desarrollo de la inteligencia artificial no sea una carrera armamentística”, de esta manera se reduce el riesgo de algún fallo posible y se desarrolla la I.A. con mayor cuidado de la seguridad.

Para la segunda etapa, Shabbir y Anwer (2015) proponen que, para el desarrollo de la inteligencia artificial de este nivel, se le incluya un entendimiento de simbolismo antes de ser utilizada, ya que de esta manera se asegura que la I.A. sea capaz de entender y seguir códigos éticos y morales.

Finalmente, para la última etapa se (que sea buena porfavor)

Conclusión

Para concluir, principalmente que la ia es una gran ayuda para la humanidad, sin embargo, hay que tener cuidado, LO DE LOS RIESGOS, PRESENTADOS CON LA SOLUCIONES

Finalmente, y luego de toda esta investigación, se puede concluir que, si al IA es desarrollada de forma segura, controlada, y responsablemente, corresponde el salto más grande de la humanidad, pero para esto debe ser investigada priorizando la seguridad y el conocimiento por sobre las ganancias posibles o armas que se puedan constituir de inteligencia artificial. De lo contrario, estamos en el camino que el empresario Elon Musk prevee en este campo, tal cual afirma en la entrevista (..) “Tamos cagaos ”

Fuentes bibliográficas

Boostrom, N. (2005). What is a Singleton? *Linguistic and Philosophical Investigations,* 5(2), 48-54. Recuperado de: https://nickbostrom.com/fut/singleton.html

Carriço, G. (2018), The EU and artificial intelligence: A human-centred perspective. *European View*, 17, 29–36. Recuperado de: http://journals.sagepub.com.pucdechile.idm.oclc.org/doi/full/10.1177/1781685818764821

Grace, K., Salvatier, j., Dafoe, A., Zhang, B. y Evans O. (2017), When Will AI Exceed Human Performance? Evidence from AI Experts. *eprint arXiv,* Recuperado de: https://arxiv.org/pdf/1705.08807.pdf

Hwang, G. (2014), Trends in extreme learning machines: A review. *Neural Networks* 61, 32-48. Recuperado de: https://www-sciencedirect-com.pucdechile.idm.oclc.org/science/article/pii/S0893608014002214

Hwang, T. (2018), Computational Power and the Social Impact of Artifical Intelligence. *eprint arXiv*. Recuperado de: https://arxiv.org/ftp/arxiv/papers/1803/1803.08971.pdf

Shabbir, J. y Anwer, T. (2018). Artificial Intelligence and its Role in Near Future. *eprint arXiv,* Recuperado de: https://arxiv.org/pdf/1804.01396.pdf

Scherer, M. (2016). Regulating Artificial Intelligence Systems: Risks, Callenges, Competencies, and Strategies. *Harvard Journal of Law y Technology,* 29, 354 – 399. Recuperado de: http://search.ebscohost.com.pucdechile.idm.oclc.org/login.aspx?direct=true&db=asn&AN=118277780&lang=es&site=ehost-live

Silver, D., Huang, A., …, Graepel, T. y Hassabis, D. (2016), Mastering the game of Go whith deep neural networks and tree search. *Nature* 529, 484-489. Recuperado de: https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf

Turchin, A. y Denkenberger, D. (2018). Clasification of global catastrophic risks connected with artificial intelligence. *AI & SOCIET.* 1-17, Recuperado de: https://link.springer.com/article/10.1007/s00146-018-0845-5

Whitworth, B. (2009), A Comparison of a Human and Computer Information Processing. *Encyclopedia of Multimedia Technology and Networking*, 2, 230 – 239. Recuperado de: http://brianwhitworth.com/braincomputer.pdf

Winter-Levy, S. y Trefethen, J. (2016), Safety First: Entering the Age of Artificial Intelligence. *World Policy Journal*, 33. Recuperado de: <https://muse.jhu.edu/article/613182>