Análisis Político No. 11

Observatorio de la Política Internacional

Universidad de Costa Rica (UCR)– Universidad Nacional (UNA) **6 de febrero de 2021**

Tatiana Peña Sequeira
Estudiante de Relaciones Internacionales
Universidad Nacional, Costa Rica

Las armas letales autónomas se han convertido en tema de mucha importancia en la actualidad, principalmente por su posible y pronta distribución. El propósito de este escrito es destacar el peligro de estas armas a partir de los sesgos que presenta la Inteligencia Artificial, específicamente en el reconocimiento facial, que podrían llegar hasta estas armas y convertirlas en una amenaza mayor a la humanidad. Para ello, en primera instancia se va a explicar qué es un arma letal autónoma, sus problemas y las críticas del derecho internacional de los derechos humanos y el derecho internacional humanitario, posteriormente cómo es que se crean los algoritmos para la inteligencia artificial y cómo estos pueden estar sesgados, por último, se tomarán en cuenta estos datos para identificar el peligro de sesgos en este tipo de armas.

Las armas letales autónomas, están siendo desarrolladas por países como los Estados Unidos, Israel, Corea del Sur, Rusia y Gran Bretaña, estas tienen como característica una gran autonomía para seleccionar sus objetivos y atacarlos a partir de sensores sin intervención humana (Campaign Stop Killer Robots, 2020). Este tipo de armas permitiría que los humanos estén físicamente ausentes en el momento del uso de la fuerza, lo que se justifica como beneficioso en cuanto a las insuficiencias que se presentan a la hora de enfrentar estrés y peligro en situaciones de riesgo (Bieri y Dickow, 2014).

No obstante, se crean otros nuevos problemas, porque primero, los humanos son excluidos de la toma de decisiones y de las consecuencias de estas (Bieri y Dickow, 2014), lo que removería la responsabilidad personal de las decisiones tomadas (Bieri y Dickow, 2014) que a su vez crea una brecha en la rendición de cuentas, porque el arma o el robot no puede ser considerado culpable, de manera que no sabe con certeza sobre quien debería caer la responsabilidad en caso de muertes o asesinatos de civiles.

Segundo, los humanos se encontrarían tanto física como emocionalmente ausentes de la situación, lo que establecería una despersonalización de la violencia y del uso de la fuerza (Bertelli, 2020) y eliminaría toda posibilidad de compasión, porque al asesinar personas desde lejos puede provocar un distanciamiento emocional de la violencia, de manera que una mayor abstracción de la situación podría agravar la tendencia (Bieri y Dickow, 2014), esto incluso podría provocar un mayor probabilidad de la guerra al no tener que arriesgar las vidas propias en una situación de conflicto (Sharkley, 2019).

Tercero, este tipo de armas no podrían cumplir con la capacidad de distinguir entre civiles y combatientes, personas heridas o que se han rendido, esto porque simplemente no se puede codificar ni programar quién o qué es un civil y que no (Bieri y Dickow, 2014 y Sharkley, 2019). Y tampoco podrían estos algoritmos entender completamente el contexto de la situación en la que se encuentran por la diversidad de situaciones que se presentan en un conflicto (Sharkley, 2019).

Por estas razones es que se han dado discusiones sobre la legalidad de estas, así como sus consecuencias en cuanto a lo deontológico, la seguridad y los avances tecnológicos (Campaign Stop Killer Robots, 2020). Todas las situaciones referidas a la toma de decisiones sobre la vida de una persona requieren juicio y entendimiento humano (Sharkley, 2019), de manera que su uso no es moralmente justificable (Bieri y Dickow, 2014).

Estos avances podrían cambiar la forma en la que se ejerce la guerra y así también el trabajo de las operaciones policiales (Bertelli, 2020),

lo que ha despertado las alertas en cuanto a los derechos humanos porque se considera que estas son incapaces de cumplir con el derecho internacional de los derechos humanos y el derecho internacional humanitario (Sharkley, 2019), debido a que atentan contra el derecho humano de la vida, de la seguridad y la dignidad humana, también viola los principios de la prohibición de la tortura y otros actos o castigos inhumanos y degradantes (Bertelli, 2020).

Esta preocupación se refleja en las negativas por varios países, grupos de expertos en la inteligencia artificial y organizaciones no gubernamentales. Además de la creación de la campaña Stop Killer Robots que ha comenzado la búsqueda y la necesidad de un tratado vinculante que las prohíba del todo o limite su funcionamiento con un control humano significativo, en el que el control al seleccionar y atacar a los objetivos sea sustantivo (Campain Stop Killer Robots, 2020), o sea que este incluido en la toma de decisiones. Esto principalmente como una prioridad humanitaria, y una necesidad tanto legal como moral (Campaign Stop Killer Robots, 2020). Costa Rica es uno de los países que participa activamente en la Campaña para su prohibición en conjunto otras 30 naciones (Human Rights Watch, 2020).

Pero ¿esto qué tiene que ver con los sesgos en el reconocimiento facial de la inteligencia artificial? Las armas autónomas letales operan mediante el reconocimiento facial, este se basa en algoritmos de aprendizaje automático que se entrenan con datos etiquetados (Boulamwini y Gebru, 2018). El reconocimiento facial funciona a través de un entrenamiento se le da al algoritmo, con una serie de imágenes de una persona, eso hace que aprenda a prestar atención a las características que son confiables para determinar si se trata del mismo individuo (Gavie, Bedoya y Frankle, 2016).

Una vez establecido el entrenamiento, se puede fijar cuáles imágenes determina mejor el algoritmo, de manera que para que el reconocimiento facial funcione debe, primero detectar las caras de las personas en las imágenes, después de esto la adecua para procesarla, posteriormente extrae las características que ha sido enseñado a cuantificar y lo compara de manera que emite una puntuación numérica en cuanto a la similitud en las características de otras imágenes en las que puede haber una coincidencia (Gavie, Bedoya y Frankle, 2016). Toda la matemática dentro de este, en el reconocimiento facial, puede incluir millones de variables que optimizan el proceso de entrenamiento, esta complejidad es lo que le da la capacidad de aprender, pero hace más difícil examinarlo y generalizar sobre su comportamiento (Gavie, Bedoya y Frankle, 2016).

Estudios actuales han demostrado que los algoritmos pueden ser entrenados con datos o información sesgada (Boulamwini y Gebru, 2018) esto porque las imágenes que se utilizan tienden a tener las mismas características, solo de una etnia por ejemplo, por lo que el algoritmo al tiempo reconocerá algunos rasgos más que otros, que podría hacer que este funcione mejor con una etnia que otra, o incluso dejar por fuera un grupo étnico completo si al algoritmo no se le entrenó con las características que les define (Gavie, Bedoya y Frankle, 2016). Esto se puede deber a que se utilizan investigaciones antiguas de reconocimiento facial, a las cuales no se le hacen cambios o también, a las propias experiencias del programador que están sesgadas lo que hace que el algoritmo se concentre en unos tipos de características y no en otras (Breland, 2017).

Estos estudios han determinado que lo sistemas de reconocimiento facial, tanto los desarrollados en Occidente como en Asia, tienden a trabajar mejor con sus respectivas poblaciones (Boulamwini y Gebru, 2018). Un estudio del 2011 identificó que los códigos creados por personas blancas en Francia, Estados Unidos y Alemania identifican mejor características caucásicas y los algoritmos desarrollados en Japón, China y Corea del Sur reconocen mejor caras con características del este de Asia (Garvie y Frankle, 2016).

Lo que sugiere que las condiciones y el contexto en donde es desarrollado un algoritmo tiene influencia en su desempeño y la precisión de sus resultados. También se considera que hay otro problema que se da al reconocer caras de personas con tez oscura, esto porque se considera que estas no pueden ser reconocidas porque los ingenieros que dominan el sector de la tecnología y quienes escriben estos algoritmos son mayormente hombres blancos (Buolamwini citada en Breland, 2017).

Lo que sugiere es que simplemente no funciona para las personas de tez morena u oscura (Breland, 2017), debido a que estos sistemas de reconocimiento facial tienden a identificar erróneamente o no identificar del todo a estas personas. Los estudios del Instituto Nacional de Estándares y Tecnología han revelado que los algoritmos tienen problemas al reconocer personas con piel oscura, estos tomaron en cuenta a más de 50 compañías y reveló que estas detectan mejor a los hombres que a las mujeres y tiene un mayor desempeño en piel clara que oscura (Simotine, 2019). Por ejemplo, los últimos algoritmos de Idemia mezcla las caras de piel más clara con piel oscura, tanto en mujeres y hombres; el de Microsoft e IBM tienden a ser perfectos en hombres de piel clara, pero falla un 20 % en cuanto a mujeres con piel oscura, el de Amazon tiende a los mismos resultados (Simotine, 2018).

A modo de conclusión se puede determinar que los sesgos se pueden dar en todos los algoritmos, porque estos son creados por personas que tienen una visión de mundo parcializada, y los problemas en el reconocimiento facial son reflejo de esos sesgos. Estos algoritmos son usados por la policía, especialmente en los Estados Unidos, por lo que estos sesgos pueden hacer que civiles inocentes sean identificados como sospechosos (Garvie y Frankle, 2016), tal ha fue el caso de Nijeer Parks, que es la tercera persona en ser arrestada erróneamente basado en una coincidencia en un reconocimiento facial inexacto (Hill, 2020). Esta situación es preocupante para el Sur global por la proyección de esta tecnología y los riesgos que conlleva para el derecho internacional de los derechos humanos y el derecho internacional humanitario, siendo necesario discutir públicamente la prohibición o regulación de las armas autónomas letales.

Referencias

Bertelli, D. (26 de junio del 2020). *An overview on human rights, artificial intelligence and autonomous weapons systems.* Ius in itinere. https://www.iusinitinere.it/an-overview-on-human-rights-artificial-intelligence-and-autonomous-weapons-systems-29017

Bieri, M y Dickow, M. (2014). Lethal Autonomous Weapons Systems: Future Challenges. CCS *Analyses in Security Policy*, 164, pp. 1-4. https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/91585/1/eth-46945-01.pdf

Breland, A. (4 de diciembre del 2017). How white engineers built racist code – and why it's dangerous for black people. *The Guardian.* https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police

Buolamwini, J. y Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, pp. 1–15, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Campaign to Stop Killer Robots (2020). *Key elements of a treaty on Fully autonomous weapon*. https://www.stopkillerrobots.org/wp-content/uploads/2020/03/FAQ-Treaty-Elements.pdf

Campaign to Stop Killer Robots (2020). *Learn.* https://www.stopkillerrobots.org/learn/

Garvie, C. y Frankle, J. (7 de abril del 2016). Facial-Recognition Software Might Have a Racial Bias. *The Atlantic*. https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-rec()ognition-systems/476991/

Garvie, C., Bedoya, A. y Frankle, J. (18 de cotubre del 2016). *The perpetual line-up unregulated police face recognition in America.* Perpetual Line-up. https://www.perpetuallineup.org/

Hill, K. (29 de diciembre del 2020). Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match. *The New York* Times. https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html#click=https://t.co/U4KGWXBjon

Human Rights Watch. (2020). *Killer Robots Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control:* Estados Unidos.

Sharkley, A. (2019). Autonomous weapons systems, killer robots and human dignity. *Ethics and Information Technology*, 21, pp.75–87 https://link.springer.com/content/pdf/10.1007/s10676-018-9494-0.pdf

Simotine, T. (22 de julio del 2019). The Best Algorithms Struggle to Recognize Black Faces Equally. *Wired*. https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/

Simotine, T. (6 de febrero del 2018). Photo Algorithms ID White Men Fine—Black Women, Not So Much. *Wired*. https://www.wired.com/story/photo-algorithms-id-white-men-fineblack-women-not-so-much/