**Violación de la privacidad a causa del empleo inmoderado de Big Data**

**Tesis:**

El implemento de Big Data en plataformas de alcance masivo para generar beneficios económicos puede poner en riesgo la privacidad y dignidad de los usuarios de internet.

**Desarrollo:**

A continuación, comenzaremos con el desarrollo del ensayo mencionando brevemente los rasgos generales del *big data* y algunos de los hechos actuales que giran en torno a la utilización de macrodatos en la economía mundial para lograr situar nuestra relación con ellos.

Antes que nada, mencionamos los usos más destacados de los macrodatos, en palabras de Bhatt, et al.:

La aplicación del *aprendizaje automático* es perfecta para explotar el conocimiento oculto dentro de un gran volumen de conjuntos de datos distintos con poca dependencia de la supervisión humana. Aprende en función de las entradas y / o salidas de datos disponibles, básicamente impulsadas por datos y se ejecuta, capaz de manejar una gran variedad de variables, así como la complejidad de los datos, que es bastante esencial en el mundo de big data actual. El *aprendizaje automático* consta de varias disciplinas de análisis de datos, desde el análisis predictivo y la minería de datos hasta el reconocimiento de patrones, y se utilizan varios algoritmos para estos fines[[1]](#footnote-2) (Bhatt, et al. 254). [Traducción propia]

Para un buen número de tareas, es cada vez más sencillo prescindir de las manos del humano, y el análisis de datos es ahora una de ellas, *big data* es capaz de encontrar información implícita de alta confiabilidad. Si nuestra información fuera un diamante, el *big data* sería el mejor pulidor en el mercado, el cual aclama poder quitar hasta la cantidad más milimétrica de carbón que impida poder observar la gema desde cada ángulo posible.

En cuanto a los hechos de actualidad. en primer lugar, es necesario entender que, aunque los macrodatos suenan a una especie de fenómeno tecnológico de una película distópica, en realidad están presentes en la vida de cualquier persona que use un dispositivo electrónico conectado a internet. De acuerdo con Wasastjerna, las empresas más importantes de la década representan papeles de importancia que aumenta en la economía actual. Compañías como Facebook, Amazon y Google han adquirido poder y alcance mundial mediante la recolección y comercio de los datos de los usuarios (419). Se tratan de empresas cuyos productos son usados diariamente por millones de personas. No es ninguna coincidencia que, estas empresas de alcance mundial influyan en la economía mundial, pero ¿cómo es que generan tanto dinero?, ¿acaso nuestras interacciones representan algún beneficio económico para estas macroempresas? La información es un producto inasible bastante valioso, el cual miles de empresas están dispuestas a adquirir, cada interacción que tenemos con internet destila una pizca de nuestra identidad.

Específicamente, esa pizca de identidad anteriormente mencionada junto con muchas otras recolectadas a lo largo del tiempo que hemos ocupado internet, directa o indirectamente construyen nuestro perfil de clientes, como Peng, Sheng-Lung, et al. advierten:

Las compañías más importantes de la década para la economía mundial implementan algoritmos de análisis de macrodatos de forma que recopilan los datos obtenidos de los usuarios de redes sociales como Instagram, Whatsapp y Facebook. A través de sus interacciones como *likes*, amigos en común e historial de búsqueda, generan publicidad con base en la personalidad, tal que maximice el alcance de las publicaciones de sus patrocinadores. (775)

Prácticamente, cada vez que le echamos un vistazo a nuestro teléfono, enviamos algunos mensajes, nos unimos a ciertos grupos, añadimos a un amigo a nuestras redes sociales o le damos un simple *like* a un meme completan los datos que estas empresas tienen acerca de nosotros, como resultado, observamos publicidad más parecida a nuestros gustos e intereses.

Siendo aún más específicos con el punto anterior, la información puede generar más información. Conforme a lo que indica O’Leary:

Utilizando datos de Twitter, los investigadores han comenzado a investigar una amplia gama de eventos, incluidos el envenenamiento de alimentos, eventos políticos, fraude, propagación de enfermedades, venta de libros y música, y otras actividades. Aunque Twitter se puede utilizar para generar estas inferencias, son claramente diferentes que los propósitos originales de los mensajes. Como resultado, podríamos preguntarnos si dichos usos son éticos y si dichos usos son una violación de la privacidad (O’Leary 92). [[2]](#footnote-3) [Traducción propia]

La interacción masiva entre usuarios de la misma región parece ser uno de los grandes motivos del surgimiento de información, la cual alude a eventos de gran impacto social. Claramente, *big data* resulta de gran utilidad para la investigación sanitaria y policiaca, incluso en un futuro cercano podría ser una de las herramientas más poderosas para combatir el crimen y tener un control de la salud de la región donde habitemos. Sin embargo, el problema principal es que la mayoría de las veces, el análisis de información no se limita exclusivamente a investigaciones de salud y policiacas, sino también empresariales o comerciales que sobrepasan la ética y privacidad.

Puede decirse que, para grandes empresas, nuestra intimidad es un negociazo para vendernos todo lo que sea posible. En palabras de Bandara, Ruwan, et al:

Los consumidores se han convertido en "generadores de datos ambulantes" en el nuevo ecosistema de mercados de datos personales donde los datos de los consumidores se recopilan, extraen y comercializan incesantemente. Este entusiasmo por los datos del consumidor aumentó notablemente con la inducción de la analítica de big data. (Bandara, Ruwan, et al. 563). [[3]](#footnote-4) [Traducción propia]

Desafortunadamente, hasta cierto punto, el ciudadano promedio se ha convertido en un número, donde las grandes empresas tienen casi todos los billetes y uno es un billete (o posiblemente un centavo) más en sus carteras. Sin duda alguna, no hay mejor negocio que saber todo acerca de tu cliente, y en caso de no ser tu tipo de cliente, referenciarlo a una empresa amiga que sí lo sea. El negocio de maquillajes, teléfonos o ropas es algo del pasado, el negocio de la información es lo de hoy.

En contraste, no todo lo relacionado con macrodatos desemboca en actos negativos. La opinión de Takle, et al. plantea lo siguiente:

Big Data, cuando se procesa adecuadamente, ha demostrado ser bastante significativo para impulsar el crecimiento de varios sectores, como sistemas de recomendación que son comúnmente vistos por locales basados en la web como Amazon, Flipkart, para asesorar productos a los consumidores de acuerdo con sus hábitos de compra son algunas de las implementaciones más notables de Big Data Analysis. Facebook sugiere "personas que quizás conozcas", lugares para visitar y recomendaciones de películas basadas en nuestros intereses (Takle, et al. 57).[[4]](#footnote-5) [Traducción propia]

Por su puesto que el uso pertinente de las tecnologías de aprendizaje automático como *big data* son tremendamente benéficas, pues permiten tener toda clase de productos a nuestro alcance, la distancia ya no es una barrera para adquirir los productos que deseemos, especias del medio oriente, figuras de colección, máquinas que limpian el suelo, etcétera. Asimismo, ahora es mucho más probable poder encontrar a aquella persona con la que platicamos, pero nunca le preguntamos su nombre, o encontrar películas de nuestro gusto sin necesidad de perder una hora tratando de descubrir si en verdad capta nuestro interés.

Desde otro punto de vista no comercial, el uso de algoritmos es una medida de seguridad estándar. Siguiendo a Mendelson, la clasificación algorítmica es confidencial entre países, la cual utiliza sensores en la vía pública, tales como cámaras de videovigilancia, bancos, escuelas y otros lugares con gran afluencia de personas, la privacidad de la información obtenida a través de estos sensores está sujeta a las políticas dispuestas por el gobierno en mando en el momento. (66) Como es lógico, la seguridad debe ser garantizada por el Estado, y, por tanto, dispositivos de videovigilancia en vías publicas resultan una herramienta útil para monitorear el orden. La medida en que estos dispositivos vigilan y monitorean depende directamente de las políticas de cada país. Sobre el mismo punto, incluso cuando no se trata de comercios entre macroempresas, los gobiernos también utilizan la tecnología con ética dudosa, tal como Shamsi y Muhammad mencionan, muchos gobiernos alrededor del mundo monitorean a sus ciudadanos, por motivos primordiales de seguridad social. Sin embargo, esta información es de suma confidencialidad y algunos de los métodos de rastreo incurren en posibles violaciones a la privacidad (75). Nadie discute que es imprescindible para los gobiernos tener información sobre sus ciudadanos, pues su trabajo es garantizar sus derechos y del mismo modo, asegurarse que cumplan sus obligaciones con el Estado. Por el contrario, existe un punto donde la vigilancia se convierte en asalto a la privacidad. Como consecuencia, la información íntima del ciudadano se encuentra expuesta, sin leyes o autoridades competentes que salvaguarden su privacidad.

En párrafos anteriores, indicamos algunos hechos actuales en torno a los macrodatos, así como su empleo en gobiernos y en empresas multinacionales, también mencionamos sus desventajas, posibles usos dudosos, aunque también sus beneficios al ciudadano. A continuación, leeremos algunas de las opiniones de los expertos en seguridad, privacidad y *big data*.

En el mercado de la información, es natural pensar que todos tienen nuestra información. No obstante Zhang Jie, et al. nos explican lo contrario:

En los últimos años, con el desarrollo y la aplicación más amplia de la tecnología de inteligencia artificial, y big data la protección de la privacidad de los datos también ha recibido cada vez más atención. El desafío que conlleva el aprendizaje automático tradicional es que, si los datos no se pueden comunicar entre instituciones, una empresa tiene una cantidad limitada de datos o unas pocas empresas gigantes monopolizan una gran cantidad de datos. (Zhang, Jie, et al. 12)

A diferencia de la opinión popular, desde el punto de vista de las empresas, los datos de las personas son demasiado valiosos monetariamente como para entregarlos a quien sea que lo necesite, como todo en una sociedad capitalista, el conocimiento tiene un costo, de tal manera que, las empresas que no aprovechan tecnologías como el aprendizaje automático (clave para el desarrollo de *big data*) que maximizan la cantidad de información acerca de posibles clientes, se encuentran en una desventaja abismal entre sus competidores.

En lo que concierne a la libertad de elegir, pueden existir ciertas limitaciones para el usuario. A juicio de Paterson y McDonagh consideremos que, “otro daño consecuente surge cuando las predicciones preventivas basadas en Big Personal Data se usan intencionalmente para disminuir el rango de opciones futuras de una persona socavando así sus libertades civiles”[[5]](#footnote-6) (8). [Traducción propia], las predicciones preventivas son de gran utilidad hasta cierto punto, dado que acortan el tiempo de toma de decisiones y ofrecen cierta seguridad de que los resultados que obtenemos apelan porcentualmente nuestros gustos e intereses, mas recorta nuestras opciones y libertad de elegir algo distinto o inusual. De alguna manera pertenecemos a un grupo arbitrario, con opciones predeterminadas y sin posibilidad de salir de ese grupo o tener si quiera la posibilidad de palpar el espectro de opciones de otras personas con diferentes características a las de uno.

Con relación al lado negativo del implemento de *big data*, como se ha referido anteriormente, los sitios recolectores de información no consideran la posible violación a ciertos derechos fundamentales. En palabras de Wasastjerna:

Hoy en día, los datos son comúnmente referidos como el "petróleo de Internet" o la "moneda de la tecnología digital economía". Si bien estas diversas analogías reflejan el valor económico de datos personales en la economía digital, no consideran que los datos también están esencialmente vinculados a la dignidad, autonomía y privacidad de las personas. La recopilación y el uso de información personal aumenta las problemáticas de privacidad. En este contexto, una cuestión emergente es el papel de leyes competentes para proteger a los consumidores de los posibles riesgos de privacidad que el creciente número de análisis de datos en los mercados digitales (418).[[6]](#footnote-7) [Traducción propia]

Claramente, de acuerdo con las palabras del autor recién mencionado, nuestra información es una parte inseparable de nuestra identidad,por ende, nuestra privacidad y datos más recónditos se ven amenazados ante cualquier empresa que esté dispuesta a pagar unos cuantos miles de dólares. A causa de éste y muchos otros problemas, los macrodatos resultan un arma de doble filo. Por un lado, actualmente es más fácil generar conocimiento acerca de una clientela en específico. Por otro lado, ese conocimiento atrae muchos riesgos para el usuario, dado que el sitio donde se encuentra el usuario tiene la obligación de defender y guardar su información personal.

Siguiendo las ideas de párrafos anteriores, la implementación del día de hoy de *big data*, tiende a limitar, y jugar con la dignidad del usuario. Similarmente, Paterson y McDonagh expresan lo siguiente:

Big Personal Data socava la dignidad humana al ignorar las elecciones de los usuarios sobre cómo se utiliza su información personal y sus sentimientos con respecto a la forma en que se procesa su información y es usada. Principalmente, socava la dignidad humana al tratar a las personas como objetos de análisis y facilitación de la toma de decisiones[[7]](#footnote-8) (7). [Traducción propia]

Parece ser que existe una línea muy delgada entre el uso óptimo de las tecnologías emergentes y su uso excesivo que minimiza la identidad humana. En suma de las opiniones anteriores, esta situación se repite en muchos aspectos de la vida cotidiana, pues es bien sabido que, el exceso de ciertas acciones, tienden a desarrollar alguna enfermedad, en este caso, la reducción de la identidad humana a simples patrones de comportamiento suele transfigurar esta tecnología novedosa en un retroceso social para el humano.

Para concluir, es justo explanar algunas de las principales restricciones y objeciones que cuestionan la tesis abarcada en este texto. Primeramente, podemos introducir algunos métodos de análisis de *big data* que pueden evitar el rastreo e investigación excesiva del usuario. Desde la posición de Shamsi y Muhammad, el empleo de técnicas de anonimización de datos con motores de búsqueda híbridos permiten controlar el flujo de datos de los usuarios (7). Es relevante observar que existen medidas de anonimización que protegen los datos del usuario, esto significa que, con los márgenes de privacidad correctos, la implementación de *big data* tiene la posibilidad de ser moderada y pertinente a los principios de la vida humana de este siglo. Por consiguiente, esta tecnología singular cuenta con pedales para seguir avanzando en esta carretera de innovación y automatización de procesos industriales, sin perder de vista elementos de mayor importancia para el humano en general.

Profundizando el párrafo anterior, existen métodos de encriptación que permiten mantener datos no necesarios como confidenciales, como la criptografía de clases públicas y cifrado de clave simétrica, las cuales transportan datos sin riesgo de que terceros puedan acceso a ellas (Takle, et al. 11). Es frecuente pensar, que la fuga de nuestra información sea un escenario real y terceros sean capaces de acceder a nuestra extensa información. Muy similar al sentimiento de invasión cuando alguien husmea entre nuestras posesiones, el conocimiento público de algunos de nuestros datos más característicos figura una tragedia para todo aquél que le preocupe su seguridad mientras navega por internet. Consecuentemente, la criptografía de clases públicas y cifrado de clave simétrica representan una fuerte medida de protección que puede ser programada en la mayoría de los sitios web que utilizamos día a día.

Análogamente, Takle et al, nos actualizan nuevamente en materia de métodos de anonimización con la idea consecuente: las aseveraciones que afirman que los macrodatos son riesgosos por la cantidad de relaciones que establecen entre datos pueden ser evitadas con la implementación de cifrado homomórfico, pues permite calcular sobre datos sin necesidad de descifrarlos, la cual es una técnica ideal para proteger la información durante el transporte y almacenado de información en los servidores (6). Así pues, una de las facetas del análisis de información, donde suele haber mayor cantidad de fugas de información, bajo los parámetros correctos puede ser más segura y confidencial mediante cifrados con claves idénticas en forma, pero con variables distintas entre sí, de esta manera, una de las mayores preocupaciones de los usuarios puede ser minimizada y su contraparte puede ser garantizada como un método de protección en el navegador.

Para finalizar, me gustaría mencionar uno de los puntos más importantes por los que los macrodatos aún deberían ser implementados en muchas áreas de investigación. Tal como postulan Harlow y Frederick, la implementación masiva de ciencias de macrodatos abre un compendio de posibilidades para descubrir patrones psicológicos individuales a través de información que, en un principio, no se clasifica con valor psicológico, pero es posible interpretar con datos pertinentes, lo cual permite entender a mayor profundidad el comportamiento humano en sistemas socioculturales. (452). Existe la oportunidad que las ciencias de macrodatos sean una forma de entender matemáticamente el cerebro humano, objeto de estudio de muchos doctores y psicólogos a lo largo de la historia, el potencial de esta tecnología podría establecer un parteaguas en la psicología social, fenómenos de comportamiento, tendencias actuales, interacciones internacionales, etcétera. Ahora, investigadores de distintas áreas del saber pueden ocupar herramientas con suficientes elementos para ser analizados junto con otros casos de estudio que permitan al humano tomar decisiones más sabias y premeditadas.

**Obras citadas:**

Bandara, Ruwan, et al. “The Privacy Paradox in the Data-Driven Marketplace: The Role of Knowledge Deficiency and Psychological Distance.” Procedia Computer Science, vol. 121, Jan. 2017, pp. 562–567. EBSCOhost, doi:10.1016/j.procs.2017.11.074.

Bhatt, Chintan, et al. Internet of Things and Big Data Technologies for Next Generation Healthcare. [Electronic Resource]. 1st ed. 2017., Springer International Publishing, 2017. EBSCOhost, search-ebscohost-com.udlap.idm.oclc.org/login.aspx?direct=true&db=cat00047a&AN=udlap.000382307&lang=es&site=eds-live.

Harlow, Lisa L., and Frederick L. Oswald. “Big Data in Psychology: Introduction to the Special Issue.” Psychological Methods, vol. 21, no. 4, Dec. 2016, pp. 447–457. EBSCOhost, doi:10.1037/met0000120.

Mendelson, A. “Security and Privacy in the Age of Big Data and Machine Learning.” Computer, vol. 52, no. 12, Dec. 2019, pp. 65–70. EBSCOhost, doi:10.1109/MC.2019.2943137.

O’Leary, D. E. “Big Data and Privacy: Emerging Issues.” *IEEE Intelligent Systems, Intelligent Systems, IEEE, IEEE Intell. Syst*, vol. 30, no. 6, Nov. 2015, pp. 92–96. *EBSCOhost*, doi:10.1109/MIS.2015.110.

Paterson Moira, and Maeve McDonagh. “Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data.” Monash University Law Review, vol. 44, no. 1, Jan. 2018, pp. 1–31. EBSCOhost, search-ebscohost-com.udlap.idm.oclc.org/login.aspx?direct=true&db=asn&AN=133727881&lang=es&site=eds-live.

Peng, Sheng-Lung, et al. Security with Intelligent Computing and Big-Data Services. [Electronic Resource]. 1st ed. 2018., Springer International Publishing, 2018. EBSCOhost, search-ebscohost-com.udlap.idm.oclc.org/login.aspx?direct=true&db=cat00047a&AN=udlap.000383120&lang=es&site=eds-live.

Shamsi, Jawwad A., and Muhammad Ali Khojaye. “Understanding Privacy Violations in Big Data Systems.” *IT Professional, IT Prof*, vol. 20, no. 3, May 2018, pp. 73–81. *EBSCOhost*, doi:10.1109/MITP.2018.032501750.

Takle, Lalita, et al. “A Survey on Data Privacy Threats and Preservation Techniques.” International Journal of Advanced Research in Computer Science, vol. 11, no. 2, Mar. 2020, pp. 57–63. EBSCOhost, doi:10.26483/ijarcs.v11i2.6522.

Wasastjerna, Maria C. “The Role of Big Data and Digital Privacy in Merger Review.” *European Competition Journal*, vol. 14, no. 2/3, July 2018, pp. 417–444. *EBSCOhost*, doi:10.1080/17441056.2018.1533364

Zhang, Jie, et al. “Edge Learning : The Enabling Technology for Distributed Big Data Analytics in the Edge.” ACM Computing Surveys (CSUR), vol. 54, no. 7, July 2021, pp. 1–36. EBSCOhost, doi:10.1145/3464419.

1. Texto original: “Machine Learning comes into the picture which is perfect for exploiting the hidden knowledge within this large volume of distinct dataset with little reliance on human direction. It learns based on available data inputs and/or outputs, basically data driven and runs at machine scale, capable of handling huge variety of variables as well as data complexity which is quite essential in today’s big data world. Machine learning consists various data analysis disciplines, starting from predictive analytics and data mining to pattern recognition and various algorithms are used for these purposes”. (Bhatt, et al. 254) [↑](#footnote-ref-2)
2. Texto original: “Using Twitter data, researchers have begun to investigate a wide range of events, including the source of food poisoning, political events, fraud, the spread of disease, sales of books and music, and other activities. Although Twitter can be used to generate these inferences, they are clearly different than the original purposes of the messages. As a result, we might ask if such uses are ethical and if such uses are a violation of privacy” (O’Leary 92) [↑](#footnote-ref-3)
3. Texto original: “Consumers have become ‘walking data generators’ in the new ecosystem of personal data markets where consumers’ data are incessantly collected, mined, and traded. This zest for consumer data notably increased with the induction of big data analytics” (Bandara, Ruwan, et al. 563). [↑](#footnote-ref-4)
4. Texto original: “Big Data when processed aptly has proven to be quite significant in boosting the growth of various sectors. Recommendation systems that are commonly seen by web based premises such as Amazon, Flipkart, to advise goods to consumers according to their buying habits are some of the most notable implementations of Big Data Analysis. Facebook suggests ‘people you may know’, spots to visit as well as film recommendations based on our interests”. (Takle, et al. 57) [↑](#footnote-ref-5)
5. Texto original: “A further consequential harm arises when pre-emptive predictions based on Big Personal Data ‘are intentionally used to diminish a person’s range of future options’,53 thereby undermining their civil liberties” (Paterson and McDonagh 8). [↑](#footnote-ref-6)
6. Texto original: “Today data is commonly referred to as the “oil of the internet” or the “currency of the digital economy”. While these various analogies reflect the economic value of personal data in the digital economy, they do not consider that personal data is also essentially linked to the dignity, autonomy and privacy of individuals. The collection and use of personal information raises privacy issues. In this context, an emerging question is the role of competition law in protecting consumers from potential privacy risks flowing from the increasing number of mergers in digital markets” (Wasastjerna 418). [↑](#footnote-ref-7)
7. Texto original: “Big Personal Data undermines human dignity by disregarding information subjects’ choices as to how their personal information is used and their feelings concerning the ways in which their information is processed and used. More fundamentally, it undermines human dignity by treating individuals as objects for analysis and facilitating decision-making, which further objectifies them” (Paterson and McDonagh 7). [↑](#footnote-ref-8)