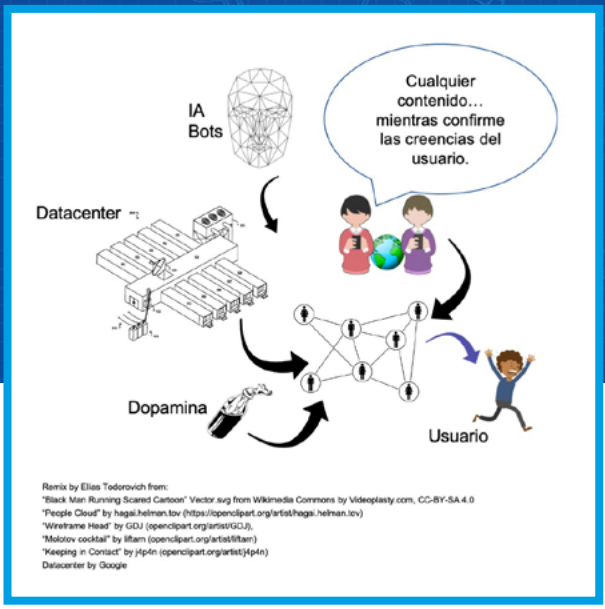
La pandemia de las redes sociales



Dr. Elías Todorovich

Dr. Elías Todorovich, Fac. de Cs. Exactas – UNICEN, Paraje Arroyo Seco S/N - Campus Universitario. Tandil, B7001BBO, Buenos Aires, Argentina. etodorov@exa.unicen.edu.ar – Facultad de Ingeniería, Universidad FASTA, Mar del Plata, Argentina.

Correo electrónico: perezjandres@hotmail.com

RESUMEN

Existen mecanismos de manipulación de la opinión pública a través de las redes sociales y otros tipos de aplicaciones informáticas, que ponen en riesgo a personas y peor aún, a la sociedad en, nada más ni nada menos que, temas de salud pública, además de economía y los valores fundamentales de nuestra civilización. Se trata del uso de tecnología muy avanzada, recientemente desarrollada, para la que todavía no se está preparado. Podría decirse que se transita un estadio del tipo “Salvaje Oeste” y se deben generar soluciones urgentemente para superarlo. La gravedad de las operaciones masivas, precisas y automáticas, que se pueden caracterizar como ciber-ataques basados en vulnerabilidades de la inteligencia humana, es clara y sus consecuencias generan alarma. Este artículo propone aplicar un conjunto de estrategias para mitigar estos riesgos y llevarlos a su mínima expresión.

ABSTRACT

We are constantly facing public opinion manipulation mechanisms through social networks and other IT applications that threaten people and, even worse, society, specifically public health issues, also economy and fundamental principles of our civilization. It is closely related to the use of advanced, recently developed technology we are not prepared to handle with. It could be said that we are going through a wild west-like stage, in which there is a need to generate urgent solutions. The severity and accuracy of these massive operations can be characterized as cyber-attacks based on weaknesses of human intelligence, and its consequences generate alarm. This article proposes applying a series of strategies to mitigate and minimize these threats.

Palabras clave: redes sociales, fake news, ciber ataques, manipulación sistematizada.

INTRODUCCIÓN

Las noticias más centrales, a nivel mundial, tienen que ver cada vez más con lo que pasa en el mundo virtual, desde escándalos hasta manipulaciones que afectan procesos electorales decisivos. Se ve el aumento de un comportamiento mucho más radicalizado, fanatizado, furioso, desde gente que echó de su casa a amigos de toda la vida y a familiares porque no tolera cuestionamientos a ideas políticas, hasta manifestaciones diversas convocadas en espacios públicos por medio de las redes sociales. ¿Qué efecto puede producir un estado de furia por un tiempo prolongado en el cuerpo y en la mente de un ser humano? Se sabe que los virus informáticos "hacen que las computadoras se hagan lentas". Eso es porque usan, para sus propios fines, una parte del tiempo del microprocesador y así la computadora llega a hacerse tan lenta que no reacciona a los comandos de los usuarios. En los humanos, la furia reduce la claridad del pensamiento, como un virus en la computadora y así, se van alejando de los intercambios de argumentos, de una escucha activa, de la búsqueda de la verdad, de cierto consenso o al menos de un respeto por el otro. Entonces, se dedican sólo a auto justificarse, cada vez más lejos de los hechos. Día a día, se hace más notoria esa vulnerabilidad que se llama sesgo de confirmación; del favoritismo se pasa al fanatismo. Y si el efecto de un virus en una computadora es producto de un ataque informático, ¿no se podrá caracterizar como un tipo de ataque el que se sufre cuando se alimenta la furia y se deteriora el discernimiento, mediante un constante flujo de desinformación y noticias [1], fotos, videos, bromas, y perfiles falsos, etc.? No se trata de recursos de manipulación nuevos, pero, el alcance, la frecuencia, la intensidad y la precisión de los estímulos no tiene precedentes. Además, cada uno de estos "servicios" tiene una tarifa. Del mismo modo, no se puede esperar menos que efectos sin precedentes. Lentamente, un nuevo monstruo fue asomando sus cabezas, que no puede pensarse, ni remotamente, sólo desde la informática, sino que requiere de un enfoque decididamente interdisciplinario.

El disparador de este artículo es la película "El dilema de las redes sociales" (TSD) [2] que presenta las tecnologías que están en los cimientos de la manipulación de la opinión pública y que se propagaron entre las redes sociales, aplicaciones de mensajería instantánea, motores de búsqueda, plataformas de contenidos, *e-commerce* y medios digitales de comunicación. En este artículo, se presentan los elementos fundamentales de este conjunto de nuevas tecnologías. Sin embargo, aunque TSD se lanzó en el 2020, sus contenidos son, mayoritariamente, de la época inmediatamente anterior a la pandemia de Covid-19. Lo que se puede anticipar de la combinación de pandemia y redes sociales es que los efectos adversos de las últimas tengan impacto en la salud pública a escala mundial, además de repercutir en la sociedad y en la economía. Para comenzar, sería muy diferente el enfoque si esta pandemia se hubiese producido con la tecnología de tan solo 10 años atrás. Hay que recordar el escenario en el que se desarrolló la gripe A (H1N1) de 2009-2010. La desinformación sobre el SARS-CoV-2 se propaga tanto o más rápido que el propio virus, especialmente al comienzo, en el 2020. Es un fenómeno que aún se está produciendo, pero se pueden ver estudios preliminares en trabajos como [3], [4].

Hay que aclarar que *machine learning* y otros algoritmos de inteligencia artificial (IA) son tecnologías extraordinariamente útiles para el desarrollo de la sociedad. En palabras de Nick Boström, "La máquina inteligente es el último invento que la humanidad necesitará realizar" [5]. Eso quiere decir que, a partir de esa inteligencia artificial general, para todos los dominios, las máquinas inventarán todo lo demás. En la misma conferencia, Boström advierte de los graves riesgos de ello. A mucha distancia de lo anterior, pero aún en el contexto de las redes sociales, cuando se indica que algo gusta o disgusta, se está entrenando al sistema para que de un mejor servicio. Sin embargo, este artículo discute el uso más nocivo de estas tecnologías, es decir, la manipulación automatizada y no los que conducen a instruir, informar, divertir y mejorar la vida de las personas. Tampoco trata sobre otros efectos dañinos como la discriminación en base a sesgos en los algoritmos de IA, ciertos usos en seguridad y justicia, como el perfilamiento criminal, detección de denuncias falsas, tecnología biométrica y procesamiento de imágenes para frenología, la cual intenta relacionar personalidad con facciones. Se visualiza que estos algoritmos requieren de grandes cantidades de datos, con un claro impacto en la privacidad, pero eso excede el alcance de este artículo, como así también el enorme consumo de energía que requiere el entrenamiento de los algoritmos de IA.

CARACTERIZACIÓN

Los contenidos que se ven en las redes sociales no son "el producto" de compañías como Facebook, Twitter, etc. Muchas veces se dijo "Cuando un producto es gratis, entonces el producto eres tú.". Los que pagan son los anunciantes; entonces ellos son los clientes. Pero, ¿qué quiere decir exactamente que el producto eres tú? En realidad, el producto es la atención de los usuarios. Pero captar atención,

¿no es capturar tiempo de vida? ¿Y cómo capto tu atención? Esa es una de las preguntas que se van a tratar de responder en este artículo.

Jaron Lanier define este punto con mayor profundidad "...el producto real es el pequeño cambio gradual e imperceptible en nuestro comportamiento. Ese es el producto, es la única posibilidad. No hay nada más en la ecuación que pudiera ser el producto... cambiar lo que haces, cómo piensas y quién eres. Es un cambio gradual, es sutil... y vale mucho dinero" (TSD, minuto 14:20).

Shoshana Zubof también va al fondo del asunto "El más grande sueño de las compañías es tener la garantía de que si ponen un anuncio, será exitoso. El negocio es vender certeza. Para eso hacen falta buenas predicciones y las grandes predicciones comienzan con un imperativo: necesitas mucha información" (TSD, 15:00).

Esa ecuación de la que hablan [6] y [7] resulta en lo que se ha dado en llamar "capitalismo de vigilancia", que obtiene beneficios con el registro y medición de toda la actividad de los usuarios, al procesarla masivamente y ofrecer servicios a anunciantes. Esto lo hacen las compañías tecnológicas que hoy están entre las más ricas de la historia conocida. Pero, lo que cierra el ciclo es predecir comportamiento en base a esos datos y maximizar el tiempo de atención, proveyendo de contenido diferenciado a cada usuario [8]. Cuando se procesa la información de los usuarios, los algoritmos de las redes sociales registran y miden todo lo que ellos hacen, durante cuánto tiempo, a qué hora del día y el lugar de cada interacción. Mediante algoritmos de IA descubren cómo te sientes, por ejemplo si te sientes solo. Descubren tu personalidad, tu estado de ánimo, tus orientaciones, tu ciclo de sueño, si usas drogas, tus vulnerabilidades. Así, estos sofisticados algoritmos construyen un modelo que predice qué video te va a gustar, qué emociones tienen un efecto en ti y en tu comportamiento.

TECNOLOGÍA PERSUASIVA

La mente humana es muy vulnerable, en otras palabras se la puede engañar con facilidad. Una demostración inocua y divertida de ello son los trucos de magia.

La tecnología persuasiva tiene como objetivo cambiar el comportamiento de personas a través de diferentes técnicas de persuasión e influencia social. Hay que referirse a estudios como los de Fogg sobre cómo cambiar el comportamiento, que se usan en diversas aplicaciones incluida la interacción humano-computadora [9], [10].

Un ejemplo de tecnología persuasiva, es el uso informático del llamado refuerzo positivo intermitente. En un momento aparecieron las interfaces gráficas donde para ver nueva información había que deslizar el dedo en la pantalla. Muy al principio, hay que recordar la incomodidad de esa especie de búsqueda secuencial, esa desazón al ir dejando atrás información no muy fácil de recuperar. Era una interacción instantánea y efímera. Los contenidos aparecen sin que se los busque, ¿cómo se hace para encontrar lo que se quiere? Pero esta manera de interactuar tiene una extraordinaria aceptación. ¿Por qué? No aparece la información que se busca, pero sí, con frecuencia, aparece información que agrada, que luego se comparte y comenta. Lejos de ser casualidad, es información seleccionada por algoritmos de recomendación con mucha precisión y presentada como si todo el mundo viera lo mismo, sin despertar la consciencia en el usuario. Eso último es clave. Es el mismo mecanismo que el de las máquinas traga-monedas de los casinos. En Psicología se llama refuerzo positivo intermitente y es lo que está detrás de ese modo de interacción tan sencillo que hasta un bebé lo puede hacer. En realidad, es tan fácil que hasta una rata lo puede hacer.

Hay otros ejemplos de manipulación y persuasión: etiquetado de fotos, activación de puntos suspensivos para retener al usuario mientras el interlocutor escribe o graba un audio.

Tristan Harris (Center for Humane Technology, humanetech.com), lo expresa del siguiente modo: "Si algo es una herramienta, de forma genuina está ahí y espera pacientemente. Si algo no es una herramienta, te exige cosas, te seduce, te manipula, quiere cosas de ti. Y pasamos de un entorno tecnológico basado en herramientas a un entorno basado en la adicción y la manipulación. Las redes sociales no son una herramienta que espera ser usada; tiene sus propias metas y sus propios medios para conseguirlas al usar tu psicología en tu contra" (TSD, -1:04:23). En tu contra quiere decir, por ejemplo, contra tu salud en el contexto de la actual pandemia. Él mismo también anticipa las consecuencias del mal uso de esta tecnología persuasiva "¿Dónde está la amenaza existencial? No se trata de que la tecnología sea una amenaza existencial, es la capacidad de la tecnología de sacar lo peor de la sociedad y lo peor de la sociedad es la amenaza existencial" (TSD, 1:17:22).

DOPAMINA

Existe una componente física, adictiva, una droga que hace que funcione esta tecnología: la dopamina. Anna Lembke lo resume así: "Las redes sociales son una droga. Estamos programados biológicamente para conectarnos con otra gente. Eso afecta directamente la producción de dopamina en

el circuito de recompensa… No hay duda de que un vehículo como las redes sociales que optimiza esta conexión entre las personas, tenga un potencial de ser adictivo" (TSD, 33:15).

Los estudios clásicos de autoestimulación cerebral de Olds y Milner mostraron que ratas con un electrodo implantado en el tracto prosencefálico medial, cuando aprendían a presionar una palanca, lograban estimular eléctricamente las neuronas sin dañarlas. Ese estímulo las llevaba a una adicción a la dopamina tal, que no se interesaban por otra cosa hasta, finalmente, morir [11]. Hoy en día, se afirma que la dopamina parece estar implicada en los estados motivacionales que nos inducen a trabajar por las recompensas, pero no en la sensación placentera que nos provocan [12]. La dopamina nos haría superar la "distancia psicológica" que nos separa de nuestro objetivo, estableciendo el límite de lo que estamos dispuestos a trabajar para conseguirlo [13].

Hay que notar que el trabajo que tenía que realizar la rata de Olds y Milner no era mucho: accionar una palanca. No tenía que resolver un laberinto o algún problema que requiriera de mucho esfuerzo para recibir dopamina. El trabajo de accionar la palanca podría considerarse más bien mínimo.

Necesitamos relacionarnos, aprobación, pertenecer a un grupo, pero alterar el proceso natural de la dopamina de la manera en que se hace, abusando de la tecnología actual, tiene graves consecuencias en varios órdenes, por ejemplo, a nivel individual, aumentando los índices de depresión y ansiedad.

Chamath Palihapitiya conecta la dopamina con las tecnologías persuasivas y sugiere efectos no sólo en las personas, sino a escala social: "Valoramos nuestra vidas según este sentido percibido de perfección porque somos recompensados con señales a corto plazo, "corazones", "me gusta"; y fusionamos eso con valores y con la verdad. Pero en realidad, es una popularidad falsa y frágil, que es a corto plazo, y que nos deja, admitámoslo, más vacíos y carentes que antes. Eso nos fuerza a entrar a un círculo vicioso donde pensamos: ¿Qué debo hacer ahora? Necesito más. Piensen en eso multiplicado por dos mil millones de personas y en cómo reacciona la gente ante la percepción de los demás. Es realmente muy malo" (TSD, 38:51).

Las técnicas de crecimiento (de número de usuarios) acelerado también están en la línea de la tecnología persuasiva.

SESGO DE CONFIRMACIÓN

En lo que hace a la polarización o grieta, Luis Chiozza, en la conferencia sobre su reciente libro [14], dice que "basta con que yo descubra un error en el razonamiento del otro, para que piense que yo pienso mejor que él, pero resulta que no estoy pensando en lo que yo pienso, pienso en el error del otro, es decir, esa es la manera estúpida de tener razón, la sinrazón del otro" y señala que esa es una reflexión inspirada por un pensamiento de José Ortega y Gasset [15]: "Pues las cosas de la política han llegado en Occidente al extremo que, de puro haber perdido todo el mundo la razón, resulta que acaban teniéndola todos. Sólo que, entonces, la razón que cada uno tiene no es la suya, sino la que el otro ha perdido". Con esto se ve que las vulnerabilidades humanas no son nuevas ni se descubrieron ayer. Lo nuevo es que se exploten masivamente con asistencia tecnología informática avanzada.

Entre los contenidos que ofrecen los algoritmos de recomendación, hay un tipo que tiene que ver con creencias y opiniones. Los algoritmos brindan argumentos a la carta para justificar cualquier postura, a gusto del consumidor y al final, sólo brindan información que refuerza opiniones. Por eso también se llaman *echo chambers* (cámaras de eco). Al algoritmo no le importa si son o no *fake news*. Además, ¿cómo se podrían distinguir unas de otras?. Quizás, podrían distinguirse justamente por el alto nivel de aceptación. El problema es si lo que se confirma mediante esta vulnerabilidad es un error. En esa respuesta emocional y exagerada que se alimenta, no interesa la búsqueda de la verdad, del conocimiento, sino, no se llamaría pulsión. Y si se parte del error, de la premisa falsa, ¿qué otra cosa se puede esperar más que la huida hacia adelante, la divergencia, las distancias cada vez mayores, la enemistad, la polarización, la alienación, la dificultad para pensar, la grieta, la falta de civismo, la desconfianza? Pero cabe recordar que en las grietas, se cae. En las huidas, alguna pared se choca y así sucesivamente. Más allá de dañar a personas, el problema escala de lo personal a lo social. Al atacar a millones aprovechando esta vulnerabilidad, hay impacto en lo público.

El sesgo de confirmación es la vulnerabilidad, las noticias falsas son el vehículo, el objetivo detrás es otra cosa diferente. Hay que recordar que, a diferencia de las noticias erróneas, las noticias falsas tienen como único objetivo manipular. Si las noticias falsas se propagan varias veces más rápido que las noticias que pueden chequearse, ¿cómo podrían ganar la carrera éstas últimas? Esa velocidad de propagación no sólo depende de usuarios humanos, sino de *bots* cuyo servicio puede comprarse [16]. Efectivamente, hay que sumar el efecto de los *bots*, que son una especie de tecnología parasitaria de la infraestructura de Internet y particularmente de las redes sociales.

POTENCIA COMPUTACIONAL

Los protagonistas de TSD opinan que no se puede luchar con este tipo de ataque desde la voluntad, ni siquiera el conocimiento detallado de todo este conjunto de tecnologías es la solución definitiva. Y eso es porque, además de la componente adictiva, se compite contra un rival cuya potencia computacional crece exponencialmente. Si se mira un solo procesador o core, desde 1978 hasta 2018, el aumento de su rendimiento fue de casi 50.000 veces [17]. ¿Cómo esperar que se puedan resolver estos problemas confiando en una simple adaptación a una tecnología con esta potencia?

Pero las computadoras más potentes se construyen usando muchos procesadores o cores, y así se tienen los procesadores *multicore*, las supercomputadoras, y se pasa a un número inmenso de cores en los *datacenters*. Chris Rowen lo sintetizaba diciendo que un core se ha convertido en el nuevo transistor [18]. Randima Fernando observa que nada de lo que tenemos ha aumentado al ritmo de la potencia de procesamiento de las computadoras. "Desde hace 60 años hasta hoy, la potencia de procesamiento ha aumentado un billón (1012) de veces… y quizás lo más importante, es que nuestra fisiología, nuestros cerebros, no han evolucionado en absoluto" (TSD, 46:00).

IA vs. yo

Tristan Harris lo expresa en otras palabras: "Los humanos no van a tener un cambio fundamental, ni a nivel mental ni físico. En el futuro, con la ingeniería genética, podríamos crear un nuevo tipo de humanos, pero siendo realistas, vivimos dentro de un sistema, de un cerebro, que tiene millones de años y tenemos una pantalla con miles de ingenieros y supercomputadoras que tiene objetivos diferentes a los nuestros. En ese juego, ¿Quién va a ganar? Díganme ustedes." (TSD, 45:25) y luego agrega "Cuando piensas en la IA, o sea una IA que acabará con el mundo, piensas en Terminator y en Arnold Schwarzenegger, ves drones y piensas que la IA matará personas, pero lo que la gente ignora es que la IA ya controla al mundo actual" (TSD, 44:55).

Los algoritmos a los que se refiere este artículo, son los de *machine learning* (aprendizaje automático). Hay que entender que en *machine learning*, los desarrolladores no programan la solución en el sentido clásico, sino que exponen a un algoritmo genérico a una etapa llamada entrenamiento, donde se le suministran grandes cantidades de ejemplos, para que pueda encontrar patrones en los datos y auto-ajustarse. Así, luego dan respuestas como las de los ejemplos con los que fue entrenado. De este modo, los algoritmos recomiendan contenidos, etc. Entonces, cuando se interactúa con las redes sociales y otras aplicaciones, no se controla el contenido que los algoritmos ofrecen, aunque aprenden de los datos. Y otras personas como los diseñadores de los algoritmos tampoco controlan eso, justamente porque estos algoritmos aprenden autónomamente. En este punto, es oportuno citar una frase atribuida a Peter Norvig, Director de Investigación de Google: "No tenemos mejores algoritmos. Sólo tenemos más datos."

Tristan Harris aclara algo importante respecto de la diferencia entre la IA de un posible futuro y la actual: "Esperamos que la IA supere las fortalezas de la inteligencia humana, que acabe con las singularidades, reemplace a los empleados y sea más inteligente que nosotros. Pero habrá un momento anterior en que la tecnología se apropiará de las debilidades humanas. Ese momento traerá adicción, polarización, radicalización, indignación colectiva, vanidad, etcétera. Eso está dominando a la naturaleza humana y es un jaque mate a la humanidad"(TSD, 53:37).

LOS EFECTOS

Se hackean elecciones [19]. La confusión y la agitación social es tal que nadie sabe lo que es verdad. No se les cree ni a los científicos de primera línea, referentes en sus disciplinas, pero sí, se cree en un mensaje que llega por Whatsapp sin la más mínima referencia, sin autoría, imposible de chequear. No se trata de la "verdad" en un sentido filosófico profundo, sino práctico. La Tierra, ¿es plana o es la esfera que se suponía que era? Puede que esa sea fácil. Pero, las vacunas contra el Covid-19, ¿son suficientemente buenas, no están suficientemente probadas o son venenosas? Políticos, ministros, cada uno de ellos en concreto ¿Es un peligrosísimo delincuente y pedófilo o un patriota y futuro prócer? No se puede ser todas esas cosas a la vez. Las diferentes miradas de una realidad, aunque muy diversas, tienen un límite. No puede ser que cualquier cosa sea verdadera.

Roger McNamee es conciso respecto de la mecánica de las *fake news*: "Con el tiempo, tienes la falsa idea de que todos están de acuerdo contigo, porque en tu fuente de noticias todos se parecen a ti. Y una vez que estás en ese estado, pueden manipularte con facilidad, de la misma forma que te manipula un mago en un truco de cartas" (TSD, 56:36).

Luis Chiozza, analizando justamente "El dilema de las redes sociales", ofrece otra mirada mucho más profunda sobre lo que antes llamamos vulnerabilidades. Chiozza evita "el juego interminable que procura identificar culpables o que se demora contemplando lo que habría que hacer y no se

hace". Cita la película "Planeta prohibido" [20], donde existe una máquina, que "yendo más allá de las intenciones con las que fue creada, no sólo se ocupa de los propósitos que habitan la consciencia, sino que también computa los deseos inconscientes que permanecen reprimidos". Y concluye que esa máquina existe: "es nuestro sistema inconsciente, que tiene acceso a la esfera motora de nuestro yo, como los actos fallidos lo demuestran. Nosotros no hacemos sólo lo que conscientemente aceptamos. Hacemos sin querer en nuestra vida, mucho de lo que inconscientemente queremos y conscientemente no asumimos." Concluye en la contratapa de su libro, "gracias a la inusitada capacidad tecnológica alcanzada no sólo logramos mucho de lo que conscientemente queremos, también estamos logrando una parte excesiva de aquello reprimido o ignorado que contiene lo peor de nuestra condición humana". Aunque con otro enfoque, en este punto se ve coincidencia con el pensamiento de Tristan Harris. A la luz de este otro enfoque, pueden surgir preguntas: ¿Cómo sabrían los algoritmos de las redes sociales que eligen contenidos, publicidad, noticias, diferenciar, ya no entre noticias chequeables y falsas, sino entre lo consciente e inconsciente de los usuarios? ¿Qué tanto aumenta el acceso del inconsciente "a la esfera motora del yo" con una exposición intensa y prolongada a estos algoritmos?

Se ve que el sesgo de confirmación es una vulnerabilidad muy notoria que no comenzó a partir del año 2000 y se explota desde tiempos más o menos remotos. Pero el amplio menú de creencias y opiniones acerca de numerosos temas no tiene precedentes en cuanto a la potencia de la realimentación, que es automática, masiva y eficaz. Tal como se señala en TSD, estos resultados sorprendieron incluso a sus jóvenes creadores. Pero una vez superado ese primer momento primordial, donde no cabrían, como propone Chiozza, culpables, esto es, la insensibilidad del mercado y la negligencia del estado, comienza una etapa de explotación de ese conocimiento. Entonces se crea un mercado donde los actores pueden comerciar atención y cambio de comportamiento. No cabe duda de que sea un negocio rentable, más en la primera época libre de regulaciones. Eso lo demuestra la migración de estas prácticas, entre las redes sociales, las aplicaciones de mensajería instantánea y otras. Ahí sí que hay culpables, los conocedores, oportunistas que aprendieron de aquellos innovadores y que hoy están entre las compañías más ricas del mundo. Después, sólo resta esperar la mayor popularización posible de la tecnología, desde usos comerciales hasta usos criminales a nivel individual y, sobre todo, social.

Guillaume Chaslot, conecta a los algoritmos de recomendación con la polarización: "Trabajé en las recomendaciones de YouTube. Me preocupa que un algoritmo en el que yo trabajé esté aumentando la polarización en la sociedad. Pero desde el punto de vista de tiempo de uso, esta polarización es totalmente eficiente para mantener a la gente conectada" (TSD, 58:40).

Si los algoritmos de recomendaciones, mediante videos, artículos, grupos de interés afines, pueden reforzar la convicción de la gente de que la Tierra es plana, ¿de qué no podrían convencerla? Y ¿A quién le puede interesar lo que un usuario opine? A la gente que vende cosas le interesa mucho, muy en particular a los que buscan lograr algún cambio en el equilibrio de poder. Los movimientos extremos ya conocían técnicas de persuasión. Lo que no había hasta hace un par de décadas eran herramientas que llegaran a amplios sectores, pero a diferencia de los medios de comunicación tradicionales, de forma personalizada, haciéndolas más efectiva. Adicionalmente, los algoritmos de recomendación hacen enfoque en una ínfima fracción de la información y llevan hacia otro sesgo, el de exposición. En palabras sencillas, es una condena a la ignorancia en el mundo digital.

Cuando una persona pasa mucho tiempo alimentando indignaciones, experimentando la ira en función de estos mecanismos de manipulación, llega un momento en el que se cambia de manera perceptible. Ya no se es uno; alguien es por uno. Si esto trasciende a escala social, la sociedad queda comprometida. Cuando la periodista Laurie Segall de CNN le preguntó a Mark Zuckerberg "Sabiendo lo que sabes ahora, ¿crees que Facebook impactó en los resultados de las elecciones de 2016?" él no lo negó y dijo: "es realmente difícil para mí tener una evaluación completa sobre eso". Pasó en todo el mundo en la última década.

Se estima que Cambridge Analytica, desde su fundación en 2013 hasta su cierre en 2018, escándalo de por medio, había influido en cientos de procesos electorales en todo el mundo. Mark Turnbull, ejecutivo de esta compañía, decía "No es bueno pelear una elección en base a hechos, porque realmente todo se trata de emociones" y Alexander Nix, CEO de la empresa, puntualizaba "son cosas que no necesitan ser verdad mientras las crean" [21]. Existen compañías sucesoras.

Si queda alguna duda sobre los peligros del uso de este conjunto de tecnologías, hay que recordar que el propio Donald Trump, siendo presidente de una de las naciones más poderosas de la Tierra, habría incitado a grupos de seguidores al asalto al Capitolio el 6 de enero de 2021 al punto que Twitter y otras redes sociales lo suspendieron

permanentemente por el "riesgo de mayor incitación a la violencia" [22]. ¿Qué supera a algo así?

Justin Rosenstein resume: "Vivimos en un mundo en el que un árbol tiene más valor económico muerto que vivo, … Este es un pensamiento cortoplacista basado en el lucro a toda costa, como si, mágicamente, cada corporación que actúa en su interés, produjera un mejor resultado… Lo aterrador, y ojalá sea la gota que rebalsa el vaso, y nos haga ver, como civilización, lo erróneo de esa teoría, es que ahora nosotros somos el árbol, somos la ballena. Pueden minar nuestra atención. Para una corporación, somos más rentables si miramos mucho una pantalla, un anuncio, que si pasamos tiempo viviendo una vida plena" (TSD, 1:24:44).

¿CÓMO SE CURA LA PANDEMIA DE LAS REDES SOCIALES?

No se puede luchar contra un conjunto de tecnologías que generan adicción únicamente desde la voluntad. Sin embargo, bajo este enfoque hay una serie de recomendaciones útiles, que incluyen desinstalar aplicaciones, desactivar notificaciones, hacer todo lo posible por mantener la privacidad de cara a los algoritmos de recomendaciones y exponernos a opiniones diferentes a las nuestras. Tener conocimiento, ayuda, pero tampoco es suficiente. La mejor opción es generar consciencia, lo cual incluye lo anterior pero va más lejos. Justamente, para algunos de los ataques descritos es clave pasar desapercibidos de un estado consciente. Desde un estado consciente se pueden observar señales, indicaciones, signos que nos permiten reconocer el problema y así, evitarlo. Por ejemplo, en el caso de las noticias, ¿figura la fuente?, ¿dice quién es el autor?, ¿está publicada en otros medios? Y lo más importante ¿te causó una reacción emocional intensa, como indignación, miedo, ira?, quizás ¿confirma tus peores sospechas? La recomendación es esperar a que disminuya la intensidad de la emoción que se generó antes de compartir o responder de alguna manera. Existen recursos más específicos, por ejemplo, se puede consultar algún sitio confiable de verificación como los que están acreditados en la *International Fact-Checking Network* (IFCN) (ver Figura 1). También se pueden usar herramientas sencillas como la búsqueda de imágenes de Google. Para niños y adolescentes hay aplicaciones y recomendaciones particulares acerca de cuándo darse de alta en redes sociales y cómo usarlas. Ese último es un asunto particularmente delicado y hay que informarse y asesorarse.

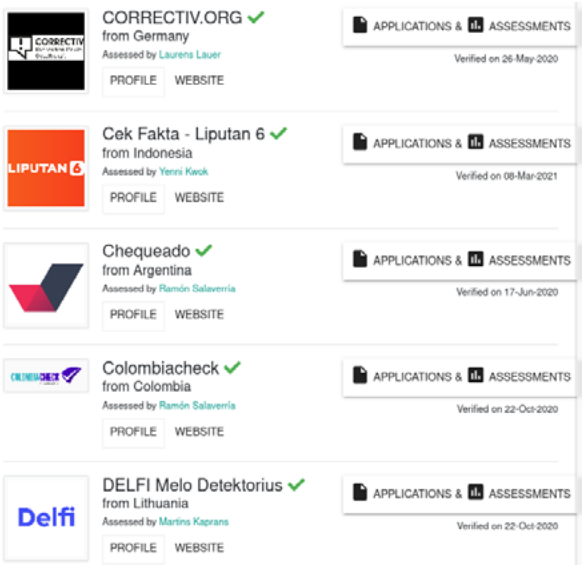


Figura 1.: Red para una gramática estándar.

Por otro lado, es creciente el número de personas cansadas de este tipo de explotación; aunque un enfoque de tipo colectivo puede tomar mucho tiempo. Igualmente, es importante comenzar cuanto antes y saber lo que sucede como usuarios de estas tecnologías. En particular, los informáticos deben tomar consciencia cabal del curso que tomó la tecnología y su impacto social, ético, político, económico y humano.

En tercer lugar, la autorregulación es inevitable por la magnitud del problema, pero Facebook, Google y otras compañías afectadas, son a la vez beneficiarias hasta límites inimaginables por este fenómeno. Lo cierto es que llevan adelante iniciativas para mitigar la difusión de las *fake news* marcando publicaciones engañosas, limitando la cantidad de reenvío de mensajes, etc. Pero, ¿cómo pueden saber entre millones de millones, cuáles noticias son chequeables y cuáles son fabricadas? Como señala [18] deberían ser más transparentes tratando de ajustar los algoritmos para favorecer la información de calidad, mitigando la difusión de noticias, especialmente por *bots*. Es más, las compañías deberían abrirse a auditorías externas. Algo positivo del lado de las compañías es la creación, en 2016, de una asociación llamada Partnership on AI, cuyo propósito es investigar buenas prácticas para el desarrollo de sistemas que usan IA. Recientemente, en noviembre de 2020, esta organización puso a disposición una base de datos sobre incidentes de IA (AIID) [23], inspirada en bases de datos similares de seguridad informática como el sistema *Common Vulnerabilities and Exposures* (CVE).

Otro enfoque sería incentivar a investigadores y empresas de ciber-seguridad a que desarrollen métodos y herramientas para detener este tipo de manipulaciones digitales, por ejemplo detec-

tando bots y perfiles falsos usando IA, o desarrollando apps que adviertan al usuario cuando los algoritmos tratan de manipularlos. Recientemente, Truepic Inc. (truepic.com) logró integrar una cámara de fotos con hardware seguro en un teléfono móvil. Así, las fotos contienen una prueba criptográfica de autenticidad, con lo cual se podrían tener este tipo de garantías de autenticidad de fotos y videos en un futuro cercano.

Otra estrategia es que surjan alternativas más éticas a las actuales en redes sociales, mensajería instantánea, motores de búsqueda, etc., con un compromiso por la protección de la privacidad, por la transparencia y el rechazo de las llamadas tecnologías persuasivas.

Respecto de la normativa, tendría que trabajarse en ello urgentemente y de manera internacional. Si bien va muy retrasada globalmente, hay que reconocer el esfuerzo de algunos países. Por ejemplo, hay que citar el Reglamento General de Protección de Datos (*General Data Protection Regulation*, GDPR) de la Unión Europea relativo a la protección de las personas físicas en lo que respecta al tratamiento y circulación de sus datos personales [24]. Sin embargo, los esfuerzos nacionales rápidamente caen en trabas jurisdiccionales: los servidores están en un país, el ataque es originado en otro, contratan servicios localizados en otro, que se pagan mediante bancos en otros, donde puede que ni siquiera haya tratados. Por otro lado, la normativa impacta en la práctica lícita y deja al resto en el camino criminal, como ocurre con cualquier delito informático.

También se propuso la creación de organismos regulatorios que aprueben - o no- el uso de IA que pueda perjudicar a personas en el presente o en el futuro, de un modo similar a como lo hace la *Food and Drug Administration* (FDA) en EEUU o la Administración Nacional de Medicamentos, Alimentos y Tecnología Médica (ANMAT) en Argentina, con los alimentos y medicamentos.

Del lado de la legalidad, se puede trabajar en legislación y consensos entre compañías, gobierno, usuarios y anunciantes. Eso es imprescindible y urgente, pero del lado ilegal no hay ley ni consenso que valga. Por eso, la propuesta es generar consciencia, comunicar efectivamente, instalar la discusión de este tema y aplicar todas las medidas para mitigar el impacto negativo de estos ataques hasta llegar a su mínima expresión, como puede ocurrir con otros ataques como el *spam*, o el *phishing*. Si bien, es verdad que el *phishing* está en el primer puesto de los incidentes de *malware* y generan ingresos en el mundo del cibercrimen, ¿cómo serían las cosas sino hubiera tantas campañas informativas y otras múltiples acciones *anti-phishing*?

Finalmente, una posible reacción a un ataque de tipo informático por la vía de vulnerabilidades de la inteligencia humana, es el de la desintoxicación digital [25], en la que intervienen profesionales y la persona interesada se abstiene, durante un tiempo, de usar teléfonos y computadoras, para dar espacio a la reflexión y toma de consciencia sobre los efectos que puede provocar este conjunto de tecnologías en los niveles de atención, privacidad, relaciones interpersonales, estrés, etcétera.

CONCLUSIONES

Cuando se desarrolla tecnología, es imperioso que se haga el máximo esfuerzo por entender las potenciales ramificaciones e implicaciones de su uso. Si hay consecuencias por el uso de la tecnología como los aquí descritos, no se puede decir "no lo sabíamos". Si llevo un teléfono inteligente a mi casa y se lo doy a un niño o adolescente, quiero tener cierta tranquilidad en que eso no le va a causar ningún daño. Los beneficios de la tecnología son indiscutibles pero hace falta responsabilidad cuando se desarrolla tecnología que usa IA. Como sociedad necesitamos que se cumplan ciertas garantías que nos brinden seguridad.

Se describieron vulnerabilidades (refuerzo positivo intermitente, sesgo de confirmación, etc.), luego, ataques efectivos que explotan esas vulnerabilidades, que lo pueden hacer de manera no dirigida, es decir, masivamente, automáticamente, sutilmente, sin que uno se de cuenta, y hasta podría decirse "a bajo costo". ¿No sería hora de pensar en eso desde un enfoque de ciber-seguridad? ¿No haría falta algún tipo de "actualización" para solucionar esas vulnerabilidades? En otras palabras, no tan del mundo informático, ¿hace falta la intervención de algún tipo de política de salud pública, de legislación o de educación?

Al igual que con el problema de la desinformación, hay que aceptar que la manipulación de la opinión pública mediante tecnología informática es un problema de seguridad. El impacto dañino en la sociedad ya es indiscutible. ¿Podría ser peor? Se puede imaginar a gente mucho menos sujeta a valores o normas que los CEOs de Silicon Valley. Efectivamente, estas tecnologías se van masificando y eso da lugar a negocios como el de las fake news, los data brokers, granjas de trolls, etc.

Durante el 2020, por la pandemia de Covid-19, se estima que la transformación digital que se iba a producir en 5 años como mínimo, ya se produjo. Esto solamente puede intensificar los desafíos aquí planteados, porque esa evolución fue demasiado acelerada. Se trata de un avance extraordinario donde no hubo otra opción que asumir todos los riesgos. En ese contexto, no se puede esperar sino un número creciente de ataques.

Como civilización, enfrentamos problemas de

tal magnitud que quizás no podamos resolverlos ni siquiera haciendo pleno uso de nuestras facultades. ¿Vamos a resolverlos, con nuestras capacidades disminuidas, hackeadas, vulneradas? Por todo lo aquí presentado, se debe trabajar urgentemente, utilizando diferentes estrategias para minimizar los riesgos de estas tecnologías.

AGRADECIMIENTOS

A María Vanesa Aranda, Hugo Javier Curti, Pilar Fernández, Liliana Suárez y Fernando Zizzias por sus comentarios y discusión de temas.

REFERENCIAS

[1] Lazer, D.M.J.; Baum, M.A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; Schudson, M.; Sloman, S.A.; Sunstein, C.R.; Thorson, E.A.; Watts, D.J.; Zittrain, J.L. (2018). The science of fake new, *Science*, 359(6380), 1094-1096.

[2] Orlowski, J.; Coombe, D.; Curtis, V. "The Social Dilemma" (2020). ("El dilema de las redes sociales") docuficción, lanzada vía Netflix.

[3] Allcott, H.; Boxell, L.; Conway, J.; Gentzkow, M.; Thaler, M.; Yang, D.Y. (2020), Polarization and Public Health: Partisan Differences in Social Distancing During the Coronavirus Pandemic. NBER Working Paper No. w26946

[4] Pennycook, G.; McPhetres, J.; Zhang, Y.; Lu, JG.; Rand, D.G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*. 31(7), 770-780.

[5] Bostrom, N. (2015) "What happens when our computers get smarter than we are?" [Video]. TED Conferences. https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are

[6] Lanier, J. (2018). *Ten Arguments for Deleting Your Social Media Accounts Right Now*, Henry Holt and Co.

[7] Zubof, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs.

[8] Bakshy, E.; Messing, S.; Adamic, L.A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science,* 348(6239), 1130–32.

[9] Fogg, B.J. (2002). *Persuasive Technology: Using Computers to Change What We Think and Do*, Morgan Kaufmann.

[10] Fogg, B.J. 2020. *Tiny Habits: The Small Changes That Change Everything*, Houghton Mifflin Harcourt.

[11] Olds, J.; Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain. *Journal of Comparative and Physiological Psychology,* 47, 419-427.

[12] Agustín-Pavón, C.; Martínez-Ricós, J.; Martínez-García, F.; Lanuza, E. (2007). Effects of dopaminergic drugs on innate pheromone-mediated reward in female mice: A new case of dopamine-independent "liking." *Behavioral Neuroscience*, 121(5), 920–932.

[13] Salamone, J.D; Correa M. (2012). The mysterious motivational functions of mesolimbic dopamine. *Neuron,* 76(3):470-85.

[14] Chiozza, L. (2020). *La peste en la colmena. Utopías y distopías en la red*.

[15] Ortega y Gasset, J. Ensimismamiento y alteración. (1939), Obras Completas, vol. 5, Alianza Editorial y Revista de Occidente. Madrid, 1983.

[16] Shao, C.; Ciampaglia, G.L.; Varol, O. et al. (2018). The spread of low-credibility content by social bots. *Nat Commun,* 9, 4787.

[17] Hennessy, J.L.; D.A. Patterson. (2018). *Computer Architecture, A Quantitative Approach*, 6th Edition, Morgan Kaufmann.

[18] "In Conversation with Tensilica CEO Chris Rowen," in IEEE Design & Test of Computers, vol. 25, no. 1, pp. 88-95, Jan.-Feb. 2008

[19] Allcott, H.; Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31 (2), 211-36.

[20] Forbidden Planet (Planeta prohibido en España; El Planeta Desconocido en Hispanoamérica), es una película de ciencia ficción estadounidense de 1956 dirigida por Fred M. Wilcox y protagonizada por Walter Pidgeon, Anne Francis y Leslie Nielsen.

[21] "Revealed: Trump's election consultants filmed saying they use bribes and sex workers to entrap politicians". Channel 4 News Investigation Team. 19 March 2018.

[22] Twitter Inc., "Permanent suspension of @realDonaldTrump", 8 January 2021. Recuperado de https://blog.twitter.com/en_us/topics/company/2020/suspension.html

[23] McGregor, S. "When AI Systems Fail: Introducing the AI Incident Database", November 18, 2020, Recuperado de https://www.partnershiponai.org/aiincidentdatabase.

[24] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[25] Schmuck, D. (2020). Does Digital Detox Work? Exploring the Role of Digital Detox Applications for Problematic Smartphone Use and Well-Being of Young Adults Using Multigroup Analysis. *Cyberpsychology behavior and social networking,* 23(8), 526-532.