MANIPULACIÓN POLÍTICA, *FAKE NEWS*, DESINFORMACIÓN Y SALUD PÚBLICA

CÓDIGO ROTO

Los secretos más peligrosos de Facebook

JEFF HORWITZ

Ariel

Jeff Horwitz

Código roto

Los secretos más peligrosos de Facebook

Traducción de Jorge Paredes

Ariel

Título original: *The Facebook Files*

Primera edición: mayo de 2024

Avda. Diagonal, 662-664, 08034 Barcelona
Editorial Ariel es un sello editorial de Planeta, S. A.
www.ariel.es
www.planetadelibros.com

ISBN: 978-84-344-3769-2
Depósito legal: B. 6.820-2024

Impreso en España

Índice

1

El regreso de Arturo Béjar al campus de Facebook en Menlo Park en 2019 fue como volver a casa. El campus era más grande que cuando se marchó en 2015 —el personal de Facebook se había ido duplicando cada año y medio—, pero el ambiente no había cambiado demasiado. Los ingenieros se desplazaban entre los edificios montados en bicicletas de la compañía, corrían por una pista de ochocientos metros que atravesaba los jardines situados en la azotea y se reunían en las cafeterías que conferían a las gigantescas oficinas de Facebook un aspecto más humano.

Béjar había vuelto porque sospechaba que algo se había atascado en Facebook. Durante los primeros años alejado de la compañía, mientras esta había recibido un aluvión de críticas negativas por parte de la prensa —que se habían acumulado como si se tratara de agua en un pozo—, había confiado en que Facebook hiciera frente a las acusaciones sobre sus productos de la mejor manera posible. Sin embargo, había empezado a notar cosas que parecían fuera de lugar, detalles que daban la impresión de que a la compañía no le importaba lo que experimentaban sus usuarios.

Béjar no podía creer que aquello fuera verdad. A punto de cumplir los cincuenta, consideraba que sus seis años en Facebook eran el punto culminante de una carrera tecnológica de lo más afortunada. A mediados de la década de 1980 era un adolescente que programaba sus propios juegos de ordenador en Ciudad de México, cuando conoció por casualidad al

cofundador de Apple, Steve Wozniak, el cual estaba estudiando español en México.

Después de un verano en el que el encandilado adolescente le hizo de guía turístico, Wozniak le dio a Béjar un ordenador Apple y un billete de avión para que visitara Silicon Valley. Los dos mantuvieron el contacto y Wozniak le pagó a Béjar la carrera de Informática en Londres.

«Tú simplemente haz algo bueno por la gente cuando puedas», le dijo Wozniak.

Tuvo éxito. Después de trabajar en una cibercomunidad visionaria, aunque condenada al fracaso, en la década de 1990, Béjar pasó más de una década como «jefe paranoico» de la, en su día, legendaria División de Seguridad de Yahoo. Mark Zuckerberg lo contrató como director de ingeniería de Facebook en 2009, tras una entrevista realizada en la cocina del propio consejero delegado.

Aunque el área de conocimiento de Béjar era la seguridad, asumió la idea de que proteger a los usuarios de Facebook significaba algo más que limitarse a mantener alejados a los delincuentes. Facebook seguía contando con cuentas de «chicos malos», pero el trabajo de ingeniería que la plataforma necesitaba tenía tanto que ver con la dinámica social como con el código.

Al principio de su ejercicio en el cargo, Sheryl Sandberg, directora de operaciones de Facebook, le pidió a Béjar que llegara al fondo de las denuncias de desnudos por parte de los usuarios, cuyo número se había disparado. Su equipo analizó las denuncias y vio que eran abrumadoramente falsas. En realidad, lo que sucedía era que los usuarios encontraban fotos suyas poco favorecedoras publicadas por amigos y trataban de que fueran eliminadas denunciándolas como pornográficas. Decirles a los usuarios que dejaran de hacerlo no sirvió de nada. Lo que hicieron fue darles la opción de denunciar que *no les gustaba* una foto en la que aparecían, describir cómo se sentían y animarlos a compartir ese sentimiento en privado con su amigo.

Las denuncias por desnudez se desplomaron aproximadamente a la mitad, recordó Béjar.

Unos cuantos éxitos de ese tipo llevaron a Béjar a crear un equipo llamado Protect and Care ('protección y cuidado'), un laboratorio de pruebas de proyectos para evitar experiencias negativas en línea, fomentar las interacciones civilizadas y ayudar a los usuarios en riesgo de suicidio, cuyo trabajo era innovador e importante. La única razón por la cual Béjar abandonó la compañía en 2015 fue que se hallaba inmerso en un proceso de divorcio y quería pasar más tiempo con sus hijos.

Aunque ya se encontraba fuera de Facebook cuando los escándalos empezaron a acumularse tras las elecciones de 2016, los seis años que Béjar pasó allí inculcaron en él un mandato que llevaba mucho tiempo incorporado al código de conducta oficial de la compañía: «Asume que la intención es buena». Cuando sus amigos le preguntaban por *fake news*, intromisión extranjera en las elecciones o datos robados, Béjar daba la cara por su antigua empresa. «La dirección ha cometido errores, pero cuando se le ha informado al respecto siempre ha hecho lo correcto», decía.

Sin embargo, a decir verdad, Béjar no pensaba tanto en las tribulaciones de Facebook. Al haberse incorporado a la empresa tres años antes de su salida a bolsa, el dinero no era un problema y se dedicaba a fotografiar la naturaleza, a realizar una serie de colaboraciones con el compositor Philip Glass y a restaurar coches con su hija Joanna, la cual, con catorce años, todavía no tenía edad para conducir. La chica documentó sus avances en la restauración de un Porsche 914 —un modelo de 1970 que era objeto de burlas por tener una estética que recordaba a una caja de pizza— en Instagram, la cual había sido adquirida por Facebook en 2012.

La cuenta de Joanna alcanzó cierta popularidad y fue entonces cuando las cosas se pusieron un poco feas. A la mayoría de sus seguidores les entusiasmaba que una chica se interesara por la restauración de coches, pero algunos hicieron gala de una misoginia repugnante, como un tipo que le dijo a Joanna que era objeto de atención «solo porque tienes tetas».

«Por favor, no hables de mis tetas menores de edad», replicó Joanna Béjar antes de denunciar el comentario en Insta-

gram. Algunos días más tarde, Instagram le notificó que la plataforma había revisado el comentario del hombre. No infringía las normas comunitarias de la red social.

Béjar, que había diseñado el procedimiento anterior al sistema de denuncias que acababa de encogerse de hombros ante el acoso sexual a su hija, le dijo a la chica que la decisión era una casualidad. Sin embargo, al cabo de unos meses, Joanna le dijo a Béjar que un chico de un instituto de una ciudad vecina le había enviado una foto de su pene a través de un mensaje directo de Instagram. Joanna le contó a su padre que la mayoría de sus amigas habían recibido fotos parecidas, y que simplemente trataban de ignorarlas.

Béjar se quedó helado. Los adolescentes que se exhibían ante las chicas eran asquerosos, pero probablemente no se sacaban la polla cuando se cruzaban con una chica en el aparcamiento del colegio o en el pasillo de una tienda. ¿Por qué Instagram se había convertido en un lugar en el que se aceptaba que aquellos chicos lo hicieran de vez en cuando, o que chicas jóvenes como su hija tuvieran que encogerse de hombros ante ello?

El antiguo Equipo de Protect and Care había sido rebautizado y reorganizado tras su marcha, pero Béjar seguía en contacto con mucha gente de Facebook. Cuando empezó a acribillar a sus antiguos colegas con preguntas acerca de la experiencia de los jóvenes en Instagram, le respondieron ofreciéndole un contrato de asesoramiento. Tal vez podría ayudarles con algunas de las cosas que le preocupaban, dedujo Béjar, o, como mínimo, responder a sus propias preguntas.

Así fue como Arturo Béjar se encontró de vuelta en la sede de Facebook. Entusiasmado y enormemente animado —la reacción de Béjar ante el aprendizaje de algo nuevo e interesante es una forma de expresar que le iba a estallar la cabeza—, tenía acceso privilegiado gracias a su familiaridad con los más altos directivos de Facebook. Autodefiniéndose como un «mexicano que va por libre», se puso a leer atentamente investigaciones internas y a organizar reuniones para debatir cómo las plataformas de la compañía podrían prestar un mejor servicio a sus usuarios.

Indudablemente, el ambiente en la empresa se había ensombrecido durante los cuatro años transcurridos. Sin embargo, Béjar se dio cuenta de que en Facebook todo el mundo era igual de listo, amable y trabajador que antes, aunque ya nadie pensara que la red social solo tenía ventajas. La sede de la compañía —con servicio de lavandería gratuito, comida por encargo, gimnasio e instalaciones recreativas y médicas— seguía siendo uno de los mejores entornos laborales del mundo. Béjar se alegraba de haber vuelto.

Aquella nostalgia explica probablemente por qué tardó varios meses en ocuparse de la que consideraba su aportación más significativa a Facebook: la modernización del sistema de denuncias de malas experiencias por parte de los usuarios.

Se trataba del mismo impulso que le había llevado a evitar organizar reuniones con algunos de sus antiguos colegas del Equipo de Protect and Care. «Creo que no quería saber», dijo.

Béjar se encontraba en casa cuando finalmente se puso a trabajar en el viejo sistema que diseñó su equipo. Los recordatorios cuidadosamente probados que él y sus colegas habían elaborado —pidiéndoles a los usuarios que informaran de sus preocupaciones, entendieran las normas de Facebook y solucionaran los desacuerdos de manera constructiva— habían desaparecido. En su lugar, ahora Facebook exigía que los usuarios alegasen una infracción concreta de las normas de la plataforma haciendo clic en una serie de ventanas emergentes. Los usuarios suficientemente motivados para completar el proceso llegaban a una pantalla final en la que se les requería que se reafirmaran en su deseo de emitir una denuncia. Si se limitaban a hacer clic en una casilla en la que ponía «Hecho», marcada como predeterminada con el color azul de Facebook, el sistema archivaba su queja sin someterla a revisión por un moderador.

Lo que Béjar no sabía entonces era que, seis meses antes, un equipo había rediseñado el sistema de denuncias de Facebook con el objetivo concreto de reducir el número de denuncias confirmadas de usuarios, de manera que la empresa no tuviera que preocuparse por ellas, liberando recursos que, de ese modo,

podrían invertirse en la formación de sus sistemas de moderación de contenidos basados en la inteligencia artificial. En una circular acerca de los intentos de mantener el coste de la moderación de los discursos de odio bajo control, un directivo reconoció que Facebook podría haberse excedido en sus intentos por mitigar el flujo de denuncias de los usuarios: «Es posible que hayamos forzado demasiado la máquina», escribió, insinuando que la compañía tal vez no quería suprimirlas de manera tan expeditiva.

La compañía declararía posteriormente que estaba intentando mejorar la calidad de las denuncias, y no eliminarlas. Sin embargo, Béjar no necesitaba ver aquella circular para apreciar que había habido mala fe. La llamativa casilla azul era suficiente. Colgó el teléfono estupefacto. Así no era como se suponía que funcionaba Facebook. ¿Cómo se iba a preocupar la plataforma por sus usuarios si no se molestaba lo bastante en escuchar qué era lo que les parecía mal?

Era cuestión de arrogancia: se daba por supuesto que los algoritmos de Facebook no necesitaban siquiera saber lo que experimentaban los usuarios para saber qué querían. Y, aunque los usuarios normales no pudieran ver lo mismo que Béjar, acabarían captando el mensaje. A personas como su hija y sus amigas les bastaría con denunciar cosas horribles unas cuantas veces para darse cuenta de que a Facebook no le importaban. Entonces dejarían de hacerlo.

Cuando Béjar volvió al campus de Facebook, seguía rodeado de personas inteligentes y serias. No podía imaginarse a ninguna de ellas optando por rediseñar las funciones de denuncia de Facebook con el fin de engañar a los usuarios para que tiraran sus quejas a la basura, pero estaba claro que eso era lo que habían hecho.

«Tardé unos cuantos meses en plantearme la pregunta correcta —dijo Béjar—: ¿Qué hacía de Facebook un lugar en el que esa clase de intentos se desvanecían de manera natural y los usuarios quedaban sometidos?»

Sin que Béjar lo supiera, muchos empleados de Facebook se habían estado haciendo preguntas parecidas. A medida que el escrutinio de las redes sociales se intensificaba tanto desde fuera como desde dentro, Facebook había acumulado una plantilla cada vez más numerosa dedicada a analizar y abordar una serie de defectos que iban saliendo a la luz. Denominado en sentido amplio como «trabajo de integridad», este proyecto se había extendido mucho más allá de la moderación de contenidos convencional. Diagnosticar y remediar los problemas de la red social requería no solo ingenieros y científicos de datos, sino también analistas de inteligencia, economistas y antropólogos. Esta nueva clase de trabajadores tecnológicos se enfrentaba tanto a adversarios externos decididos a aprovechar las redes sociales para lograr sus propios fines como a altos ejecutivos que creían que el uso de Facebook era, en general, un bien absoluto. Cuando sucedían cosas feas en la red social homónima de la compañía, dichos directivos lo atribuían a los defectos de la humanidad.

Los miembros del personal responsables de abordar los problemas de Facebook no podían permitirse ese lujo. Su trabajo exigía entender cómo la empresa podía distorsionar la conducta de sus usuarios y cómo, en ocasiones, se «optimizaba» de forma que previsiblemente terminaría causando daños. Los trabajadores encargados de mantener la integridad de Facebook se convirtieron en los guardianes del conocimiento que el mundo exterior ignoraba que existían y a los que sus jefes se negaban a creer.

Mientras un pequeño ejército de investigadores con doctorados en Ciencia de Datos, Economía Conductual y Aprendizaje Automático averiguaba cómo su empleador alteraba las interacciones humanas, yo lidiaba con cuestiones mucho más básicas sobre el funcionamiento de Facebook. Me había mudado recientemente a la Costa Oeste para cubrir la información relativa a Facebook para *The Wall Street Journal*, un trabajo que llevaba aparejada la desagradable necesidad de fingir que escribía con autoridad sobre una empresa a la que no entendía.

No obstante, había una razón por la cual quería cubrir la información de la red social. Tras cuatro años dedicándome al

periodismo de investigación en Washington, el trabajo de información política que realizaba se me antojaba carente de sentido. Ahora, el nuevo ecosistema estaba dominado por las redes sociales, y los artículos no generaban interés a menos que fueran del agrado de los devotos de internet. Había mucha información falsa que se estaba viralizando, pero las comprobaciones de datos que yo escribía no parecían tanto una medida correctiva como un débil intento de capear las consecuencias de las mentiras.

Cubrir la información sobre Facebook era, por tanto, una capitulación. El sistema de divulgación de información y de creación de consenso del que yo formaba parte se hallaba en las últimas, así que más me valía que me pagasen por escribir sobre el que lo iba a sustituir.

Lo sorprendente fue lo difícil que me resultaba entender incluso lo más básico. Los encargados de explicar públicamente el algoritmo del News Feed —el código que determinaba qué publicaciones se mostraban a miles de millones de usuarios— recurrían a frases como «Te conectamos con lo que más importa». (Posteriormente supe que existía una razón por la cual la empresa pasaba por alto los detalles: grupos de muestreo habían llegado a la conclusión de que las explicaciones en profundidad sobre el News Feed confundían e inquietaban a los usuarios; cuanto más pensaba la gente en externalizar «quién y qué importa más» a Facebook, menos cómodos se sentían.)

En un guiño a su inmenso poder e influencia social, la compañía creó un blog llamado *Hard Questions* ('preguntas difíciles') en 2017, declarando en su entrada inaugural que se tomaba «muy en serio nuestra responsabilidad —y trascendencia— por nuestro impacto e influencia». Sin embargo, *Hard Questions* no entró nunca en detalles y, tras un par de años complicados sometido al escrutinio público, el proyecto fue abandonado discretamente.

Cuando empecé a cubrir la información sobre Facebook, la reticencia de la compañía a responder a las preguntas de los periodistas también había aumentado. Su Departamento de

Prensa—un numeroso equipo de casi cuatrocientos empleados— tenía fama de amable, profesional... y reacio a responder preguntas. Yo disponía de muchos contactos de relaciones públicas, pero no conocía a nadie que quisiera explicarme cómo funcionaban las recomendaciones de «Personas que quizás conozcas», qué señales hacían que las publicaciones polémicas se hicieran virales o a qué se refería la empresa cuando decía que había impuesto medidas extraordinarias de seguridad en medio de la limpieza étnica de Myanmar. Las recomendaciones de contenidos de la plataforma determinaban qué bromas, nuevas historias y cotilleos se volvían virales en todo el mundo. ¿Cómo funcionaba esa caja negra?

La consiguiente frustración explica cómo me convertí en un fan de cualquiera que estuviera mínimamente familiarizado con la mecánica de Facebook. Los antiguos empleados que accedieron a hablar conmigo me dijeron cosas inquietantes desde el principio. Los sistemas de ejecución automatizada de Facebook eran absolutamente incapaces de funcionar según lo previsto. Los intentos de impulsar el crecimiento habían favorecido de forma involuntaria el fanatismo político. Y la compañía sabía mucho más de lo que decía sobre las consecuencias negativas del uso de las redes sociales.

Era un tema espeluznante, mucho más cautivador que las eternas alegaciones de que la plataforma censuraba publicaciones o favorecía injustamente al presidente Trump. Sin embargo, mis fuentes de antiguos empleados de Facebook no podían ofrecerme demasiado en lo que a pruebas se refería. Una vez abandonaban la empresa, dejaban su trabajo tras los muros de Facebook.

Hice todo lo posible por utilizar a los empleados actuales como fuente, enviando cientos de notas que se reducían a dos preguntas: ¿Cómo funciona realmente una empresa que influye en miles de millones de personas? Y ¿por qué con tanta frecuencia parece que no lo hace?

Por supuesto, otros periodistas actuaron también de manera parecida. Y, de vez en cuando, obteníamos documentos aislados que indicaban que los poderes y los problemas de Facebook

eran mayores de lo que se dejaba entrever. Tuve la suerte de estar presente cuando el goteo de información se convirtió en un diluvio. Algunas semanas después de las elecciones de 2020, Frances Haugen, una directora de producto de nivel intermedio del Equipo de Integridad Cívica de Facebook, respondió a uno de mis mensajes de LinkedIn. La gente tenía que entender qué estaba pasando en Facebook, dijo, y había estado tomando algunas notas que creía que serían útiles para explicárselo.

Haugen no se atrevía a decir nada más por LinkedIn ni por teléfono, así que quedamos en una ruta de senderismo en las colinas, detrás de Oakland, aquel mismo fin de semana. Después de un paseo de medio kilómetro por los bosques costeros de secuoyas de California, salimos del sendero para hablar en privado.

Haugen fue una fuente inusual desde el principio. Las plataformas de Facebook erosionaban la fe en la sanidad pública, favorecían la demagogia autoritaria y trataban a los usuarios como un recurso explotable, declaró en nuestro primer encuentro. En lugar de admitir sus problemas, la empresa estaba introduciendo sus productos en mercados remotos y pobres donde, según ella, era prácticamente seguro que provocarían algún daño.

Dado que Facebook no se estaba ocupando de corregir sus defectos, dijo, creía que podría jugar un papel importante haciéndolos públicos.

Ninguno de nosotros tenía una imaginación lo suficientemente grandiosa como para adivinar lo que generaría aquel proyecto: decenas de miles de documentos confidenciales que demostrarían la profundidad y el alcance del daño causado a todo el mundo, desde chicas adolescentes a las víctimas de cárteles mexicanos. El alboroto sumiría a Facebook en una crisis de varios meses, y fomentaría que el Congreso, los reguladores europeos y los propios usuarios se cuestionaran el papel de Facebook en un mundo que parecía abocarse a un caos cada vez mayor.

No todas las personas con información privilegiada con las que hablaría a lo largo de los dos años siguientes compartían el

diagnóstico exacto de Haugen de lo que había salido mal en Facebook, ni su receta para remediarlo. Sin embargo, en su mayor parte estaban de acuerdo, no solo con sus antiguos colegas tránsfugas, sino también con las evaluaciones escritas de empleados que *nunca* hablaban públicamente. En los documentos internos recopilados por Haugen, así como en cientos más que me fueron entregados después de su salida, miembros de la plantilla documentaban los demonios del diseño de Facebook y trazaban planes para dominarlos. Luego, cuando su empleador no tomaba medidas al respecto, veían cómo tenía lugar una crisis tras otra, todas previsibles.

Indicara lo que indicara el manual del empleado, cada vez resultaba más difícil asumir que la intención de la empresa era buena.