COMPARECENCIA DE LA DIRECTORA DE POLÍTICAS PÚBLICAS, EMEA, DE TWITTER, SINÉAD MCSWEENEY, Y DE LA DIRECTORA DE SEGURIDAD DE TWITTER, PATRICIA CARTES, EN LA SESIÓN DE LA PONENCIA CONJUNTA DE ESTUDIO SOBRE LOS RIESGOS DERIVADOS DEL USO DE LA RED POR PARTE DE LOS MENORES EL DÍA 5 DE MAYO DE 2014

La señora **DIRECTORA DE POLÍTICAS PÚBLICAS, EMEA, DE TWITTER** (Sinéad McSweeney): Maybe I'll just say a few words of introduction first. My name is Sinead McSweeney, as the chairman said. I look after public policy for twitter in the near region which includes Europe, the Middle East and Africa. Up until very, very recently it was just me. One person looking after that region which is one of the reasons why unfortunately it has taken me such a long time to come to Spain, which is unfortunate, not just because we have many users here in Spain and we like to ensure that safety messages etc. are available wherever we have lots of users, but also because I like Spain. It would have been much nicer to have been here sooner.

So in terms of my role, I look after government relations, relationships with regulators the beginning of relationships with police to ensure that they are aware of our policies about how to acquire information from us which may be helpful with an investigation and I deal with our privacy issues and other regulatory issues.

We have a very strong and committed trust and safety team. Up until very recently, Patricia was part of our trust and safety team but I have been very fortunate to welcome her to the public policy team and she has many years of expertise and experience although you wouldn’t think so when you look at her in the area of safety particularly online safety for young people and she is now heading up our global safety outreach so but we are

fortunate that we have her in Europe. So we will get more of her time and then the rest of the world.

I think one thing that is important to understand is that very often people say Google/ Facebook / Twitter as if it was one word or as if we were all the same company or the same size. The reality is that while the word Twitter is known far and wide, the company itself is still quite small and its growth as a company in terms of actual human beings trying to keep the company going has only happened, really in the last 12 to 18 months, so I often draw the parallel that Google has more employees in Dublin than Twitter has in the whole world, so that might give you some sense of how small we are and the efforts we are making to catch up with the platform.

Patricia is going to take you through our safety policy. She will touch on the size of the platform but mainly our safety policies and safety procedures and the work that she and colleagues from trust and safety do. Hopefully that will cover a lot of the questions that you are likely to have but between us we will take care of any outstanding issues once she has gone through the presentation.

La señora **DIRECTORA DE SEGURIDAD DE TWITTER** (Patricia Cartes): Sin más dilación, comenzaré. Y hoy quería empezar intentando ilustrar un poco las dimensiones de Twitter. Y como bien decía Sinéad ahora, Twitter es una plataforma en crecimiento, si bien la empresa es mucho más pequeña, y siempre estamos intentando seguir la estela de la plataforma, pero a veces es difícil.

En cuanto a visitantes únicos, tenemos 400 millones al mes; de esos 400 millones, 230 suelen entrar de forma muy regular, eso quiere decir que entran a visitar el sitio por lo menos de forma diaria. Y vemos 1.000

millones de tuits cada dos días. Para que se den una idea de lo que esto significa, tardamos tres años, dos meses y un día en ver el primer millón de tuits, y sin embargo ahora vemos un millón de tuits cada dos días. Por lo que cuando nos enfrentamos a abuso en la plataforma, muchas veces cuando llegamos a la denuncia de abuso ese tuit ya no es importante, porque ha habido muchos tuits que le han seguido. Entonces, es primordial que intentemos dar soporte al usuario inmediatamente cuando ocurre cualquier problema en el sitio web.

La página web está disponible en 35 idiomas, tanto en la versión web como en la aplicación. Eso quiere decir que un usuario francés puede acceder a twitter.com y ver la interfaz en francés, y también puede contactar con los equipos de soporte al usuario en su idioma nativo y recibir asistencia en el idioma nativo.

Y en cuanto a empleados, tenemos unos 2.000 empleados a nivel global. Como decía Sinéad, por ejemplo Google tiene 3.000 solo en la oficina de Dublín, que es su sede europea, y nosotros tenemos 2.000 en todo el mundo, por lo que se pueden imaginar que son muchos menos empleados de los que tenemos en cuanto a usuarios, por lo que nos toca hacer a todos mucho trabajo día a día.

Para los que estén aquí que no utilicen Twitter, esto será útil; para los que utilicen Twitter, lo siento, igual se aburren ahora un momento conmigo, pero quería captar lo que es la anatomía de un tuit. Todas las cuentas en Twitter tienen un nombre que va seguido del símbolo de la arroba. Por ejemplo, aquí estamos mostrando el de uno de nuestros colaboradores en el campo de la seguridad en España, que es la asociación PantallasAmigas. Entonces, @pantallasamigas es el nombre del usuario, que les identifica, y para conectar con ellos lo único que hay que hacer es escribir @ más ese nombre, para poder conectar con ellos directamente.

El tuit tiene 140 caracteres, o menos, y es la forma que la gente tiene de comunicarse con sus seguidores. Se pueden imaginar: 140 caracteres es muy poco espacio, por lo que cuando recibimos denuncias de abuso a veces nos falta mucho contexto; es muy difícil para nosotros juzgar qué es realmente lo que está pasando, por lo que siempre les pedimos a los usuarios que nos den cuanto más contexto mejor a la hora de realizar las denuncias de abuso.

También tenemos la verificación, que es esa marca azul que se ve al lado del nombre de PantallasAmigas, en este caso, y la verificación muestra cuando hemos marcado una cuenta como auténtica en la plataforma. Y eso nos sirve para que los usuarios sepan que están interactuando con la persona o la entidad a la que realmente siguen. Por ejemplo, si yo soy fan, no sé, de Pep Guardiola, quiero asegurarme de que estoy interactuando con Pep Guardiola, y no con, igual, un fan o alguien que está haciendo una parodia de Pep Guardiola. Y la forma de saber que realmente él es con quien estoy interactuando es buscar ese símbolo azul al lado del nombre del usuario.

Y finalmente tenemos la forma de interactuar con un tuit, que está al final del tuit, en el que podemos responder, podemos retuitear, y lo que hace retuitear es simplemente mostrar ese mismo contenido en mi cuenta; es una forma de compartir el mismo contenido en mi cuenta. De forma que si yo ahora quisiera compartir este mensaje en mi cuenta, lo único que tendría que hacer es clicar en retuitear para que apareciera en la cuenta de Patricia Cartes.

También puedo marcar un tuit como favorito, y si le doy a los tres botones de más –lo veremos más adelante– tendré las opciones de denuncia de abuso, que es de lo que se encarga mi equipo. Y eso enlaza bastante bien con el equipo de Trust & Safety. No hemos conseguido traducir el nombre al castellano, porque en realidad el campo de “safety” en castellano se suele

traducir como “seguridad”, pero implica algo mucho más allá de la seguridad; no estamos hablando solo de recuperación de contraseñas o de cuentas comprometidas y haqueadas, estamos hablando de la protección al menor en un sistema de una forma más amplia.

Entonces, este equipo está especializado, hay diferentes áreas de especialización. Y entre ellas quería recalcar la de propiedad intelectual e identidad. Como decía antes, tenemos mecanismos para asegurarnos de que las cuentas que tenemos en el sitio son reales, y cuando nos encontramos con usurpación de marcas o suplantación de identidad es importante que tomemos acción.

También tenemos el equipo que se encarga de llevar los derechos de usuario y privacidad, que son temas como menores de 13 años que estén en la plataforma, derechos de imagen (alguien que sube una foto mía a Twitter y yo no quiero que esa foto esté en Twitter, cómo puedo denunciarlo y qué pasos vamos a seguir para eliminar esa imagen). Lo mismo con cualquier otro tipo de elemento de privacidad, como direcciones, números de teléfono que sean privados, etcétera.

Las políticas para anunciantes: este equipo se encarga de ver qué productos hay que restringir en qué países (por ejemplo, los anuncios de alcohol en según qué países nórdicos están prohibidos en línea), de forma que se encargan de realizar todo el estudio de mercado para todos los mercados en los que facilitemos anuncios a nuestros usuarios.

El equipo de seguridad de usuario, que es con el que yo trabajo de forma más cercana, se encarga del abuso y el acoso a un nivel bastante general, cosas como amenazas, el lenguaje de incitación al odio, autolesiones, contenido que tenga relación con el suicidio; miran todas las denuncias que recibimos sobre esas áreas y se encargan de responder a los usuarios y de atenderles.

Y finalmente, el equipo que se encarga de las solicitudes legales: ellos se encargan día a día de las relaciones con fuerzas de seguridad (en el caso de España, Guardia Civil, Policía Nacional y todas las policías regionales). Reciben las solicitudes de información, cuándo ha ocurrido un crimen, o si se quiere saber más del usuario. Y también se encargan de bloquear el contenido a nivel geográfico. Es algo que podemos hacer. Imaginémonos un contenido que sea ilegal, por ejemplo, en Turquía: podemos bloquear ese contenido para el territorio turco pero dejarlo en el sitio para el resto del mundo. Entonces, se encargan mucho de mirar este tipo de solicitudes.

Y finalmente, en la explotación de menores se encarga junto con el equipo de seguridad del usuario de mirar cualquier tipo de denuncia que nos llegue de este tipo de contenido, pero también de nuestros esfuerzos proactivos a la hora de combatir cualquier tipo de contenido relacionado con la explotación al menor. Y sobre esto voy a hablar un poco con más detalle en unos minutos.

Hay otro equipo, que es el de atención al usuario, que complementa a este primer equipo del que hablaba. Y entre estos dos equipos, que están situados tanto en Dublín como en San Francisco, podemos hacer soporte al usuario 24 horas al día, 7 horas a la semana, de forma global. Por lo que, cuando Dublín se va a dormir, San Francisco toma el relevo y se encarga de seguir con las denuncias de abuso que nos llegan.

El equipo de atención al usuario hace un soporte más genérico: se encargan del centro de ayuda, mirar cualquier tipo de errores del sistema que estemos viendo de forma sistemática. También se encargan de ver las etiquetas, que son las palabras claves que van precedidas del símbolo del sostenido, y que permiten a los usuarios agrupar conversaciones por un tema. Por ejemplo, si yo quisiera hablar del partido del Real Madrid ayer podría utilizar ese símbolo más “partido del Real Madrid”, y empezaría a

conectar con cualquier usuario que esté hablando de ese tema. Es una forma realmente de agrupar las conversaciones a través de palabras clave. Pero, como se pueden imaginar, en ocasiones se dan caso de abuso con etiquetas; puede haber etiquetas abusivas e incluso ilegales, y es algo con lo que tenemos que tener mucho cuidado.

También se encargan de *spam*, *phishing*, *malware*, toda la parte técnica de cuando una cuenta es vulnerada, e intentar restablecer esa cuenta, el cambio de contraseñas y asegurarse de que ningún usuario que haya interactuado con esa cuenta también haya sido infectado por los virus que sean.

Fotos y multimedia: hay mucho contenido multimedia en Twitter. Twitter te permite subir imágenes al sitio, y este equipo se encarga de asegurarse de que esas imágenes son legales, pero también de que no haya abuso en general con las imágenes.

Acceso a cuentas, cuentas haqueadas, restablecimiento de contraseñas, cualquier persona que tenga un problema de inicio de sesión, que tal vez ha perdido la dirección de correo asociada a su cuenta, etcétera, este equipo se encarga de restablecer todas esas cuentas.

Y finalmente, el equipo de traducción, que se asegura de que el sitio esté disponible en esos 35 idiomas de los que hablábamos antes, tanto el centro de ayuda como el sitio web como las aplicaciones a las que se puede acceder a través del teléfono o del iPad.

Hay dos formas de denunciar abuso en la plataforma, porque hemos hablado de los equipos que se encargan de mirar estas denuncias, y la forma que tiene el público de dar con ese equipo es a través de dos mecanismos principales.

El primero es el centro de ayuda, que contiene muchísima información; información muy simple, como cómo cambio el nombre de

mi cuenta o cómo subo una foto a la cuenta, a información más compleja, como cómo denuncio terrorismo en Twitter.

El segundo mecanismo de denuncia es el tuit. A nivel de tuit también hemos implementado un mecanismo de denuncia. Como decía antes, en el botón de "Más", cuando clicamos en "Más", y esto ya sea desde el sitio web, desde el teléfono móvil o desde la aplicación, tengo la opción de bloquear o reportar. Es importante marcar –esto es lo que se ve una vez que clico en bloquear o reportar–, recalcar la opción de bloquear, porque queremos que el usuario también sea capaz de protegerse a sí mismo. Y una de las recomendaciones que compartimos es siempre, si estás interactuando con alguien o alguien interactúa contigo de forma abusiva, es muy importante bloquear y no interactuar con ellos, porque muchas veces, cuanto más se interactúa, es como echar leña al fuego.

Pero aparte de la opción de bloquear, que te permite acabar con esa interactuación de forma inmediata, también puedes denunciar el tuit en particular. Y hay varias opciones: puedes denunciar el tuit como *spam*, puedes denunciar la cuenta por estar comprometida si lo que estás viendo son tuits de *spam*, y esto ocurre muchas veces, que igual ves a un amigo que está compartiendo tuits de adelgazar diez kilos en tres días, y toda la cuenta tiene ese tipo de tuits, lo más lógico es que hayan sido haqueados. Y entonces, es importante denunciar la cuenta como comprometida para que nosotros podamos restablecerla.

Y la última opción, que es realmente la opción que nos ocupa aquí, es la de cuando un usuario es ofensivo. Y cuando clicas en este enlace te vamos a preguntar exactamente qué tipo de violación de nuestras reglas estás observando. ¿Estamos hablando de acoso o estamos hablando de incitación al odio o estamos hablando de una violación de la privacidad de alguien? Y como decía antes, como 140 caracteres es muy poco espacio, siempre le pedimos al usuario que nos dé, cuanto más contexto, mejor. Y

dependiendo de la opción que elijas, eso le irá a un equipo o a otro. Y tenemos la gran suerte de tener especialistas en los diferentes equipos. Por ejemplo, el equipo de privacidad está conformado de gente que realmente ha trabajado en el campo de la privacidad y de la protección de datos durante muchos años, y son capaces de ver el tuit, saben cuáles son las reglas de Twitter, cuál es la legalidad vigente en el país en el que se encuentra el usuario y tomar la acción apropiada.

Pero podemos hablar más de eso ahora, porque hay un par de reglas de las que realmente quería hablar hoy. La primera, la de abuso, acoso y amenazas, y es importante decir que el abuso y el acoso están prohibidos en Twitter, y siempre que nos llegan denuncias de este tipo, lo que hacemos es evaluar la situación. Muchas de las interactuaciones que vemos en el sitio, la gran mayoría de las interactuaciones que vemos en el sitio web son positivas. Pero de vez en cuando tienes la situación de dos amigos, que igual se han peleado, y uno empieza a tuitear al otro de forma abusiva. Y en estos casos nos viene muy bien poder mandarle advertencias al usuario y compartir con ellos las reglas de Twitter. Y vemos que en la gran mayoría de los casos los usuarios reaccionan muy bien a estas advertencias y suelen rectificar el comportamiento que han estado mostrando en la plataforma.

Hay casos más severos en los que se crean cuentas solo para abusar de alguien, y está claro porque cuando te metes en la cuenta ves la arroba y el nombre de alguien que de forma muy consistente se ve que está haciendo abuso. Esas cuentas, obviamente, no las queremos en el sitio, y en esos casos tenemos que suspender la cuenta de forma permanente y asegurarnos de que ese usuario no está creando otras cuentas para continuar ese abuso. Y para asegurarnos de que eso no ocurre mantenemos un diálogo constante con el usuario y le vamos preguntando si ha observado en alguna cuenta nueva, si han recibido contacto nuevo por parte de ese usuario desde cualquier otro canal, para poder tomar acción.

Y en el medio también tenemos las cuentas que tal vez son abusivas, pero no es el único objetivo de la cuenta; puede ser una cuenta que estaba bien y por el motivo que sea se ponen a abusar de alguien, y en esos casos podemos hacer una suspensión temporal, y avisar al usuario, y tienen que leerse las reglas antes de ser restablecidos en el sitio web.

Entonces, tenemos muchas acciones disponibles a la hora de enfrentarnos al abuso.

El caso de las amenazas ya es más complicado. No permitimos amenazas directas, es obvio eso. Y siempre estamos mirando cuál es la posibilidad de que una amenaza en Twitter cause daño en el mundo real. Y si hay realmente elementos que nos preocupan, y creemos que la integridad de alguien está en peligro, no solo tomaremos acción en el contenido, sino que también le recomendaremos a la persona que contacte con las fuerzas de seguridad, y entonces nosotros continuaremos esa conversación con las fuerzas de seguridad. Entonces, digamos, si a mí me ha amenazado hoy alguien y hay suficientes elementos en el tuit, como por ejemplo el lugar de la amenaza, en armas que se van a utilizar, quién va a participar en el ataque y demás, y yo voy a la Guardia Civil, nosotros podemos luego mantener la conversación con la Guardia Civil para realmente llegar al meollo de la cuestión, por así decirlo.

Ha habido últimamente muchas conversaciones sobre la explotación infantil en línea, y quería aclararlo también hoy: tenemos tolerancia cero para el contenido de explotación al menor y trabajamos de acuerdo con las leyes, sobre todo las americanas, que claramente nos marcan que tenemos que denunciar cualquier tipo de contenido de este nivel a NCMEC (que es el National Center for Missing and Exploited Children). NCMEC se encarga de trabajar con las diferentes fuerzas de seguridad en los diferentes países, de diseminar la información necesaria para que la persona que haya compartido este tipo de contenidos sea llevada a la justicia. En España

trabajamos mucho sobre todo con la Guardia Civil, también con Policía Nacional y las fuerzas de seguridad regionales, pero siempre es a través de NCMEC. Y en particular la Guardia Civil tiene una conexión VPN con NCMEC, que es la forma que tienen de obtener los datos específicos de un usuario que esté compartiendo este tipo de contenido.

Este tipo de contenido es complicado. En Twitter, al ser una plataforma abierta, vemos menos de este tipo de contenido que otras redes sociales, porque no hay mucho sitio por el que esconderte, no hay muchos lugares en los que te puedas esconder. Y lo que es obvio es que los explotadores de menores no suelen compartir este contenido de forma abierta.

Aun así, es importante que seamos proactivos, porque si bien la comunidad denuncia este tipo de contenido, entre ellos los explotadores no lo van a denunciar, por lo que no podemos confiar solo en la comunidad para denunciar este tipo de tuits.

Por eso tenemos una tecnología que se llama Photo DNA, que fue desarrollada por Microsoft con Dartmouth College y es usada por todas las empresas de este tipo de industria (Facebook, Microsoft, Yahoo, Google, etcétera), que es una base de datos de imágenes que están escaneadas; funciona por un *hash*, y los *hashes* de las imágenes son un poco como el ADN de una imagen. Entonces, cuando tienes el mapa de ese ADN de la imagen, cuando una imagen es subida a un sitio web, si hay una imagen que le corresponde en la base de datos nos alerta inmediatamente, la cuesta es suspendida y le pasamos la información a NCMEC. O sea, es una tecnología proactiva que siempre está funcionando por detrás de la plataforma, y nos permite a veces llegar a los casos antes de las denuncias.

Y otra regla de Twitter que quería subrayar aquí hoy es la de suplantación de identidad. Porque, si bien consideramos cuentas de parodia, y hay muchas en el sitio, hay muchas cuentas, incluso por

humoristas famosos, que hacen cuentas de parodia sobre políticos y gente de la vida pública, no permitimos la suplantación de identidad, y es algo que siempre estamos mirando, cuál es el objetivo de la cuenta y si el objetivo es engañar al usuario o hacerse pasar por alguien, ahí es cuando tenemos que eliminar el contenido, eliminar la cuenta y ver qué acciones tomar.

Pero hay un mecanismo específico de denuncia. Como mostraba antes, cuando denuncias a un usuario como usuario abusivo, también te preguntamos si lo que estás denunciando es una suplantación de identidad, en cuyo caso vamos a mirar elementos como el nombre, la imagen, pero también si los tuits están siendo producidos en primera persona y si hay una intención paródica o no. Y a veces la línea entre la parodia y el abuso es muy delgada. Entonces, es algo que llevamos juntos el equipo de propiedad intelectual y el equipo de abusos, porque en los casos en que, sobre todo a personajes privados, la suplantación de identidad es abusiva (que es lo que suele pasar, no suplantas a alguien por la broma, sobre todo cuando son personajes privados), en ese caso tenemos que suspender la cuenta del usuario y advertirles de que están en violación de las reglas de Twitter.

Estas son las que había elegido que pensé que serían más relevantes para este grupo, pero ahora tenemos tiempo y Sinéad y yo vamos a responder cualquier pregunta que tengan, si ha quedado algo que no ha sido claro.