

¿Contenido o datos generados por el usuario?

Las dinámicas (in)voluntarias de generación de contenido en línea¹

User-generated content or data?

The (in)voluntary dynamics of online content generation

Conteúdo ou dados gerados pelo usuário?

As dinâmicas (in)voluntárias da geração de conteúdo on-line

DOI: <http://doi.org/10.18861/ic.2025.20.1.3954>

► MARCELO SANTOS

marcelo.santos@udp.cl - Santiago de Chile - Universidad Diego Portales, Chile.

ORCID: <https://orcid.org/0000-0002-2658-3764>

ORCID: <https://orcid.org/0000-0003-3580-2073>

CÓMO CITAR: Santos, M. & Román Aponte, L. (2025). ¿Contenido o datos generados por el usuario? Las dinámicas (in)voluntarias de generación de contenido en línea. *InMediaciones de la Comunicación*, 20(1). DOI: <http://doi.org/10.18861/ic.2025.20.1.3954>

► LUJÁN ROMÁN APONTE

muroman@uc.cl - Santiago de Chile - Pontificia Universidad Católica, Chile.

Fecha de recepción: 7 de diciembre de 2024

Fecha de aceptación: 6 de mayo de 2025

¹ El primer autor, Marcelo Santos, recibió financiamiento de la Agencia Nacional de Investigación y Desarrollo (ANID) a través del Fondo Nacional de Desarrollo Científico y Tecnológico, N° 11230980 y del Fondo de Financiamiento de Centros de Investigación en Áreas Prioritarias, N° NCS2024_007. La segunda autora, Luján Román Aponte, recibió financiamiento de la Agencia Nacional de Investigación y Desarrollo (ANID) a través de la Beca Doctorado Nacional, 2020, N° 21200327.

RESUMEN

La creciente dataficação y la generación involuntaria de datos simplemente por interactuar, con distintos grados de conciencia, con la diversidad de plataformas sociales digitales, exige una revisión del concepto Contenido Generado por el Usuario (CGU). En este sentido, se propone el concepto de Datos Generados por el Usuario (DGU), basado en una idea frecuentemente usada pero escasamente discutida en la literatura académica, y se aborda de qué manera los rastros digitales dejados por los usuarios son capturados y procesados por terceros. A continuación, se realiza una crítica que distingue y delimita las nociones de DGU y la definición de CGU, sosteniendo que constituyen dos ramas de investigación distintas a partir de las cuales se puede pensar la generación de contenido en línea.

PALABRAS-CLAVE: *dataficação, redes sociales, Contenido Generado por Usuario, datos generados por el usuario, privacidad.*

ABSTRACT

The increasing datafication and the involuntary generation of data –simply by interacting, with varying degrees of awareness, with a wide range of digital social platforms– calls for a reassessment of the concept of User-Generated Content (UGC). In this regard, the notion of User-Generated Data (UGD) is proposed, based on an idea that is frequently invoked but rarely discussed in academic literature. The discussion

focuses on how users' digital traces are captured and processed by third parties. A critical analysis is then offered, distinguishing and delineating the concepts of UGD and the definition of UGC, arguing that they represent two distinct research strands from which online content generation can be approached.

KEYWORDS: *datafication, social media, user-generated content, user-generated data, privacy.*

RESUMO

A crescente dataficação da sociedade e a geração involuntária de dados –apenas por meio da interação, com diferentes níveis de consciência, com diversas plataformas sociais digitais– exige uma revisão do conceito de Conteúdo Gerado pelo Usuário (CGU). Nesse contexto, propõe-se o conceito de Dados Gerados pelo Usuário (DGU), baseado em uma noção frequentemente mencionada, mas pouco debatida na literatura acadêmica. Discute-se de que forma os rastros digitais deixados pelos usuários são capturados e processados por terceiros. Em seguida, apresenta-se uma análise crítica que distingue e delimita os conceitos de DGU e CGU, sustentando que se tratam de dois campos de pesquisa distintos a partir dos quais se pode pensar a geração de conteúdo online.

PALAVRAS-CHAVE: *dataficação, redes sociais, conteúdo gerado por usuário, dados gerados pelo usuário, privacidade.*

1. INTRODUCCIÓN

La introducción de tecnologías digitales que permiten, por un lado, la creación de contenido en diversos formatos (fotos, textos, etc.), y que, a su vez, se articulan con otros componentes que posibilitan el intercambio a través de canales propios y autónomos (Castells, 2009; Santos, 2022), ha precipitado una producción de contenido independiente sin precedentes en la historia de los medios. Sin embargo, casi la totalidad de lo publicado está, literalmente, “contenido” en plataformas privadas comerciales que tienen, en alguna medida, fines de lucro. Este factor resulta relevante y tiene efectos significativos en los procesos de apropiación de datos más allá del contenido originalmente publicado o de la intención del usuario al realizar la publicación. Hipotetizamos, por lo tanto, que emerge una laguna conceptual en el campo que abarca contenidos y significados extraídos sin consentimiento, sin conciencia o intención por parte de los usuarios en dichas plataformas. Para dar cuenta de esta laguna, proponemos la conceptualización y operacionalización del término Datos Generados por el Usuario (DGU) como un concepto necesario para sustentar toda una línea de investigación crítica en plataformas digitales, para situarse en un ecosistema mediático que excede ampliamente el contenido conscientemente generado por sus autores.

En el año 2022 se conmemoraron los 20 años de la introducción del término Contenido Generado por el Usuario (CGU, o UGC en su acrónimo en inglés) al vocabulario de los medios y del campo académico. Con el objetivo de establecer una definición adecuada que permitiera articular diferentes campos en la órbita de dicho concepto, se propuso en aquel momento:

Contenido Generado por el Usuario es cualquier tipo de texto, dato o acción realizada por usuarios de sistemas digitales en línea, publicado y difundido por el mismo usuario a través de canales independientes, que incurre en un efecto expresivo o comunicativo ya sea de manera individual o combinado con otras contribuciones de la misma u otras fuentes. (Santos, 2022, p. 108)

Durante estas últimas décadas, el CGU ha estado en el centro de prácticas comunicativas de usuarios ordinarios (Chouliaraki, 2010) no profesionales, como el periodismo ciudadano, el testimonio de eventos extraordinarios (Santos, 2023), entre otros. Detrás de la popularización de los teléfonos inteligentes, hay factores económicos, sociales y culturales (Wunsch-Vincent & Vickery, 2007), pero también avances tecnológicos que han permitido dichos cambios: primero, se incorporaron funciones multimedia, lo que permitió la grabación, y, entonces, los dispositivos se volvieron realmente inteligentes y conectados de forma casi permanente a redes telemáticas. A partir de los registros aficionados pixelados de eventos extraordinarios como los atentados de Londres de 2005 (Allan, 2007; Reading, 2009) y el tsunami de Tailandia de

2006 (Wardle, Dubberley & Brown, 2014), el contenido creado por el teléfono se convirtió gradualmente en una verdadera alternativa para los periodistas, especialmente cuando el lugar sobre el que se informaba era lejano o peligroso (Wardle & Derakhshan, 2017). La producción y distribución de contenido generado por usuarios creció a un ritmo exponencial, y el manejo de los aportes de ciudadanos como fuentes de noticias se volvió en un problema cada vez más acuciante para las redacciones. Esto llevó a la British Broadcasting Corporation (BBC), por ejemplo, a implementar en 2005 lo que llamaron “BBC UGC Hub” (Williams, Wardle & Wahl-Jorgensen, 2011) para lidiar con el volumen de contenido ciudadano que se les dirigía a través de canales digitales; no solo para separar el trigo de la paja, sino para verificar la confiabilidad del contenido seleccionado.

En los inicios de los 2000, la percepción de muchos analistas sobre la creación de contenido de forma autónoma y distribuida por ciudadanos ordinarios era optimista. Por ejemplo, se argumentaba que los ciudadanos con teléfonos inteligentes supuestamente se convertían en *pequeños hermanos* (Chadwick & Howard, 2009), en alusión a la figura del “Gran Hermano”, o que construirían un mosaico de fragmentos de contenido generado por el usuario que potencialmente podría convertirse en un panóptico invertido (de abajo hacia arriba), en el que las redes sociales se volverían un arma de la ciudadanía y guardiana de la institucionalidad, incluyendo los medios tradicionales y la violencia policial (Santos, 2023).

La posibilidad de crear y publicar contenido con autonomía (Castells, 2009) se asoció, en cierta medida, con la visión normativa de que el ejercicio de la ciudadanía equivalía a participar en estos nuevos entornos digitales florecientes como ciudadanos *mediactivos* (Gillmor, 2010). Dicha percepción, esta especie de visión tecno-optimista, ignoraba que la gran mayoría de los usuarios no crean contenido, solo usan, de diversas formas, contenido creado por otros (van Dijck, 2009). Luego, dicha perspectiva fue contrastada con evidencia sobre la escasa eficacia del CGU como narrativa alternativa ante eventos extraordinarios, como las protestas callejeras, demostrado por Santos (2023) en su análisis de Twitter (hoy X). Finalmente, se produjo un giro radical de una percepción inicial de pureza/positividad cívica en las múltiples formas de participación ciudadana a través de plataformas digitales, se pasó a reconocer diversas formas de “participación oscura” que se apoderaron de buena parte de las interacciones en las redes sociales, como bien operacionalizó Quandt (2018).

A lo largo del siglo XXI, la creciente datificación de la vida ha traído consigo nuevas formas de asignar significado a los problemas sociales, funcionando como herramienta para impulsar la formulación de políticas más eficientes. En otras palabras, diversas dimensiones de la vida son gestionadas, o bien sus procesos de decisión, por datos generados en el contexto de la actividad digital de

la ciudadanía. Por su parte, Milan y Treré (2021) subrayan los riesgos asociados al aumento de la disparidad social y la explotación ciudadana, derivados de las paradojas intrínsecas de la ausencia de datos sobre las poblaciones vulnerables (Peng, 2024). En ese contexto, sugerimos analizar profundamente no solo la creación de contenido, sino también la producción de significado que se da en las plataformas digitales, o mediado por ellas. Y no solo en lo que respecta a la creación de contenido por parte de los usuarios que crean esas formas de contenido, sino también a los significados inesperados, incluso no autorizados, producidos por terceros al apropiarse de dicho contenido o sus metadatos asociados. Diferentes formas de recopilación, extracción, visualización y análisis de datos no exentos de críticas han surgido o se han vuelto más prominentes (Milan & Treré, 2021; Pellegrino, Söderberg & Milan, 2019; Rieder et al., 2015), a medida que nuestra sociabilidad se convirtió en un bien valioso del siglo XXI (van Dijck, 2013).

No estamos exentos, las y los académicos, de tales críticas, ya que somos parte del sistema que se dedica a enmarcar los problemas de nuestro tiempo. No solo muchos hemos sido parte de la narrativa tecno-optimista, sino que el entusiasmo y la curiosidad científica pudo haber impulsado prácticas de *extractivismo de datos* con un sentido ético poco atento a las especificidades del mundo digital (Herrada, Santos & Barbosa, 2024). Sin embargo, también recae sobre la labor académica la responsabilidad de seguir trabajando estos problemas, incluso si ello implica contradecir reflexiones formuladas en el pasado.

Demostraremos en este artículo que el término CGU ha encapsulado dos campos diferentes. A medida que avancemos en una distinción más detallada entre ambos, quedará claro que hay líneas de investigación que resuenan en cada uno de ellos. Este artículo es, por tanto, una reflexión crítica sobre la literatura, pero también sobre nuestra propia conceptualización previa de CGU (Santos, 2022). El resultado es su desmembramiento en un concepto doble: Contenido Generado por el Usuario y Datos Generados por el Usuario, dos objetos de estudio y dos campos.

Este artículo desafía la idea detrás de la parte “contenido” del acrónimo y la problematiza en ese contexto, afirmando que hay factores que se alinean mejor con el concepto de “datos”. Para llegar a esta distinción, primero haremos una breve arqueología del CGU y cómo este se convirtió en una mercancía; luego discutiremos la distinción entre contenido y dato, para finalmente circunscribir los campos para CGU y DGU.

2. MERCANTILIZACIÓN DEL CGU

Aunque la literatura de algunas disciplinas ha descuidado la operacionalización del contenido de las redes sociales como contenido generado por el usuario (CGU), hemos afirmado anteriormente que el CGU es el elemento

sociotécnico que se encuentra en el epicentro de las redes sociales (Santos, 2022). Redes sociales como X, Facebook, Instagram o TikTok ganan sentido debido a la contribución no solo de actores destacados como celebridades, instituciones, etc., sino también de los usuarios comunes y corrientes. Mientras que los primeros muchas veces están profesionalizados², los segundos alimentan dichas plataformas con contenido, motivados por algún grado de propósito comunicativo, como una mayor visibilidad, la creación de redes, entre otros. En ese sentido, el CGU se asemeja al funcionamiento del mundo *fandom*, en cuanto a la generación y distribución de contenidos generados por “personas que se reúnen en grupos por preferencias e intereses en común” (Torti & Schandor, 2013, p. 3), que son aprovechados por la industria del entretenimiento para obtener un mayor rédito económico. Dicho proceso es similar al que sucede desde las últimas décadas con los datos de los usuarios, transformados en mercancía, ya que la sociabilidad se ha convertido en producto (van Dijck, 2013; Srnicek, 2017) en un contexto económico en el que se ha prestado mucha atención al extractivismo de datos.

Dicha práctica generalizada del extractivismo de datos por parte de plataformas de redes sociales, agencias políticas, de *marketing*, entre otras, y su progresiva normalización (Segura & Waisbord, 2019) ha sido comúnmente identificada como parte del proceso de *datificación de la sociedad* (Milán & Treré, 2021), enmarcado en una economía caracterizada como *capitalismo de plataforma* (Srnicek, 2017). En esta economía datificada, solo unas pocas plataformas poseen el mercado debido, entre otras prácticas comerciales, al efecto red, un fenómeno que permite que las plataformas se vuelvan más valiosas gracias a la interacción masiva de los usuarios que utilizan las herramientas y funcionalidades que brindan los entornos digitales. El efecto tiene muy poco que ver con los valores ampliamente defendidos del libre mercado: mientras más actividades realizan los usuarios en las plataformas, estas empresas refuerzan su monopolio mediante la centralización de datos dentro de su entorno (Srnicek, 2017). Como muestra de lo anterior, al explorar los negocios, asociaciones e integraciones de socios de las 20 redes sociales más utilizadas, van der Vlist y Helmond (2021) advirtieron la existencia de un ecosistema de plataformas integrado en lugar de una plataforma única, donde la gobernanza y el control son ejercidos a través de acuerdos de asociación e infraestructura con intermediarios de datos que mapean las huellas digitales de los individuos, profundizando el proceso de capitalización de las plataformas. Para empeorar las cosas, muchas veces, estos entornos extremadamente ricos en datos son reutilizados con fines que no fueron declarados en un principio, entrando en áreas de regulación gris y ética objetable.

Las empresas Meta son quizás el ejemplo más claro. El escándalo de

2 Usuarios con mucha visibilidad usan profesionales de comunicación y marketing y community managers profesionales

Cambridge Analytica en 2016 (Cadwalladr & Graham-Harrison, 2018) fue una muestra flagrante de la falta de control de los usuarios sobre su privacidad a través de una práctica cuestionable de elaboración de perfiles psicológicos y microsegmentación de propaganda política a través de Facebook (Risso, 2018). Al mismo tiempo, fue una excusa para que la empresa cerrara su API, limitando aún más la posibilidad de que observadores independientes como periodistas o académicos pudieran monitorear, estudiar y escrutar su actividad (Bruns, 2019). Unos años más tarde, Meta compró WhatsApp por la extraordinaria suma de 19 mil millones de dólares (Olson, 2014), con el compromiso de mantener algunas características estratégicas en la aplicación, por ejemplo, su ausencia de publicidad y sus estándares de privacidad. Sin embargo, poco después de las promesas, Meta comenzó a invertir en la plataforma de WhatsApp, comprometiendo la privacidad de los usuarios de muchas maneras, desde compartir sus metadatos con Facebook para mejorar su atractivo comercial hasta permitir que las empresas recopilen datos de los usuarios a través del entorno empresarial de WhatsApp (Johns, Matamoros-Fernández & Baulch, 2024). Las campañas #blocksidewalk y *Fuck Off Google* impulsadas por activistas de datos como formas de resistencia a la mercantilización de sus datos a partir de su apropiación por parte de las grandes potencias tecnológicas a través de prácticas poco éticas (Charitsis & Laamanen, 2024), advierten la necesidad de mirar más allá de la funcionalidad de las tecnologías digitales como herramientas de organización de las iniciativas de los movimientos sociales, como por ejemplo la Primavera Árabe, para buscar otras alternativas dentro del mismo capitalismo de datos. Ambientes supuestamente protegidos de aplicaciones cifradas pueden crear condiciones de activismo seguro como el “activismo detrás de escena” propuesto por Treré (2020) o la articulación del disenso en ambientes autoritarios como en Malasia (Johns, 2020) o Rusia (Santos, Saldaña & Tsyganova, 2024). Sin embargo, muchas veces la percepción de seguridad y privacidad de los usuarios no corresponde con la que efectivamente ofrecen dichas plataformas (Herrada, Santos & Barbosa, 2024), por lo que se sugiere que estos adapten sus comportamientos en entornos digitales de acuerdo a sus percepciones de confianza y privacidad en cada ecosistema digital, para sentirse más seguros (Saura, Palacios-Marqués & Ribeiro-Soriano, 2023).

En este mundo de Big Data, surgen múltiples estrategias de recolección y análisis de datos, mientras la sociedad en su conjunto intenta ponerse al día en términos normativos, regulatorios e incluso éticos. A medida que avanza la carrera, surgen muchas áreas grises, donde las plataformas privadas han estado recopilando una gran cantidad de datos para convertirlos en productos, mientras que las plataformas estatales los recopilan no solo para desarrollar políticas públicas, sino también para encuestar y ejercer control sobre su ciudadanía. Ejemplos de esto último son el impulso para que la plataforma de mensajería

estatal Soroush en Irán sustituya a Telegram (Kargar & McManamen, 2018), o aplicaciones populares en China como Douyin, la variante local de TikTok y WeiXin, la versión doméstica de WeChat, que presentan variaciones en lo que respecta a los estándares de privacidad y protección de datos en sus versiones nacionales e internacionales (Jia & Ruan, 2020). Las empresas chinas que ejecutan las aplicaciones están obligadas a adherirse a cumplir con dichos estándares en relación al contenido que pueden publicar, por lo que invierten en sistemas de filtrado de contenido y curaduría humana (Ryan, Fritz & Impiombato, 2020).

A medida que pasamos de formas intencionadas, hasta ingenuas, de producción de contenido por parte de usuarios que buscan, en última instancia, impulsar alguna forma de sociabilidad, hacia procedimientos opacos de recolección de datos resignificados con escasa o nula conciencia por parte de quienes los generan, se observa un desplazamiento del *contenido* generado por el usuario a los *datos* generados por el usuario. Desde la perspectiva del usuario, se argumentará en la siguiente sección que “dato” alude al significado construido por terceros, muchas veces a costa de los usuarios que lo generan, mientras que “contenido” remite a los datos creados inexorablemente unidos al contexto, lo que les confiere un significado u objetivo en alguna medida previstos por los usuarios que los generan.

3. DE CGU A DGU

Si bien algunos contenidos encajan fácilmente en la categoría de CGU, como las imágenes o textos creados y publicados por los usuarios a través de sus propios canales (Santos, 2022), como lo son las cuentas individuales de Twitter/X (Santos, 2023), otros tipos de información no reflejan claramente la intención de compartir o crear contenido por parte de los usuarios. La dimensión principal de esta discusión reside en la tensión en la tensión entre, por un lado, una creación de contenido que conlleva cierta conciencia e intención por parte del usuario, y por otro, una producción de información que se basa en la recontextualización, agregación y resignificación del CGU, dando lugar a nuevos contenidos orientados a audiencias distintas de aquellas originalmente previstas. En otras palabras, ¿qué contenido está *generando* el usuario y qué contenido está *siendo generado* a sus espaldas?

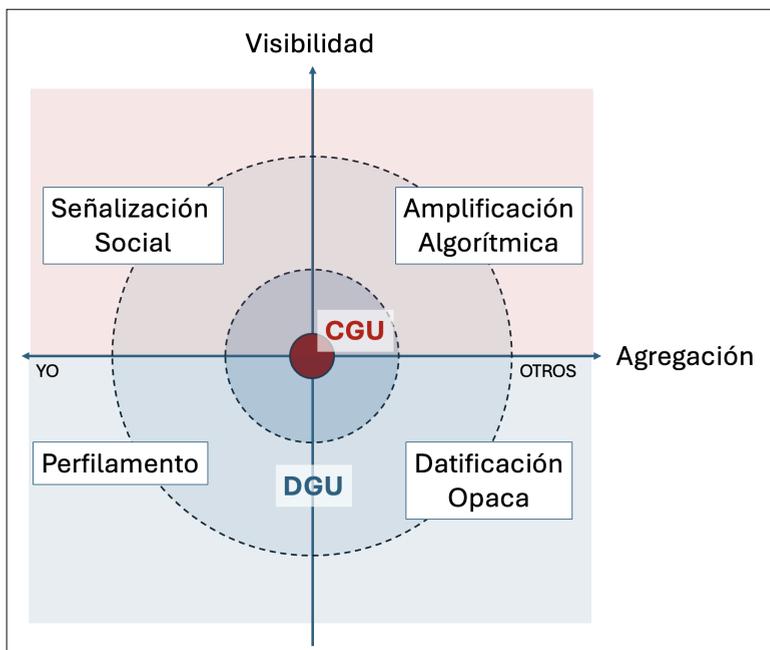
Las pocas definiciones de DGU presentes en la literatura suelen estar condicionadas por la disciplina de origen del estudio, como la definición de Saura, Ribeiro-Soriano y Palacios-Marqués (2021) en el campo del *marketing* y la innovación: “DGU incluye todas las formas de información y datos que los usuarios generan individualmente como resultado de la interacción con los elementos que componen cualquier mercado digital (acciones, experiencias, sentimientos, comentarios, reseñas, etc.)” (p. 1). Los autores luego se

propusieron discutir los límites de la privacidad en la apropiación de DGU en iniciativas impulsadas por la innovación, pero su definición no se distingue adecuadamente de las definiciones comunes de CGU, y la limitan a los “mercados digitales”. Un ejemplo de definición de CGU que no se distingue claramente de la definición de DGU de Saura, Ribeiro-Soriano y Palacios-Marqués es la de Kaplan y Haenlein (2010): “El contenido generado por el usuario (CGU) puede verse como la suma de todas las formas en que las personas hacen uso de las redes sociales” (p. 61). Otra definición de contenido creado por usuario que se menciona con frecuencia se refiere a: “1) contenido que se pone a disposición del público en Internet, 2) que refleja una cierta cantidad de esfuerzo creativo, y 3) que se crea fuera de las rutinas y prácticas profesionales” (Wunsch-Vincent & Vickery, 2007, p. 9). En esta definición, los datos generados por el usuario no son más que un subconjunto poco delimitado del contenido generado por el usuario.

Pero, ¿cómo distinguir *datos* de *contenido*? En las ciencias de la información, se establece una distinción clave entre datos e información: mientras que la última se define como datos contextualizados, los primeros representan la forma bruta de cualquier entrada digital registrada en una base de datos (Setzer, 2004). En este marco, *la información* está estrechamente relacionada con lo que consideramos contenido, mientras que los datos no necesariamente conservan su vínculo con el contexto original o pueden referirse a un contexto diferente del original sin (o con muy poca) conciencia o intención por parte del o los usuarios que los generan.

Como tal, los datos generados por el usuario se refieren a aquellos generados a partir de la actividad en plataformas digitales que adquieren significado a medida que son procesados de manera distinta a la originalmente prevista (extraídos, recontextualizados, agregados, etc.), utilizados para generar nuevos significados, con poca o ninguna conciencia por parte de los usuarios. Por lo tanto, mientras que el CGU es originalmente generado por usuarios comunes, se publica en sus propios canales individuales y su trayectoria es visible para su autor, el DGU desarticula el significado original del contenido, vulnerando principios de privacidad, intención y conciencia. El dato del DGU, por lo tanto, se transforma en información en la medida en que es recontextualizado y resignificado por los extractores de datos. En otras palabras, mientras que el CGU implica conciencia de la visibilidad y control del contexto del contenido generado, el DGU ignora ambas dimensiones, lo que lleva a diferentes escenarios, representados en el Gráfico 1. En el centro está el CGU original. A medida que actores externos, como otros usuarios, algoritmos, agencias de *marketing*, extractivistas de datos, etc., actúan sobre él, el contenido puede fluctuar a través de las diferentes dimensiones, tal como se ilustra en lo graficado.

Gráfico 1. Mapa de posibilidades que distinguen el contenido generado por el usuario (CGU) de los datos generados por el usuario (DGU) en dos dimensiones: visibilidad y agregación. Mientras que el contenido visible que conserva su significado original se considera CGU, el que se reutiliza mediante medios opacos de procesamiento se considera DGU



Fuente: Elaboración de los autores.

Cuando un contenido generado por el usuario se viraliza “orgánicamente” con la ayuda de algoritmos y otros datos del usuario (por ejemplo, “me gusta”, “compartir”, etc.), mantiene en gran medida su significado original y el usuario puede observar cómo circula incluso si dicha visibilidad se ve amplificada por factores externos. Esta exposición puede traer consecuencias tanto positivas como negativas, y aunque no siempre son consecuencias necesariamente negativas, ciertamente no son deseadas de forma deliberada. A medida que el contenido acumula “amplificación algorítmica”, el usuario pierde el control sobre el contenido. Otro tipo de agregación de lo que originalmente era CGU es la recolección acumulativa de datos como fuentes para la minería de datos a gran escala, de formas que son invisibles para el usuario. En estos casos, el significado previsto del contenido se pierde y su uso, análisis e interpretación quedan opacos para la mayoría de la sociedad. Por otra parte, cuando la agregación se realiza a nivel del usuario, la cara visible de los datos agregados es algo así como el currículum de las redes sociales, es decir, las señales sociales que informan la percepción del usuario dentro de una comunidad. La versión

opaca de dicha agregación es la elaboración de perfiles invisibles (y por lo tanto carentes de conciencia) de los usuarios, como lo ejemplifica el escándalo de Cambridge Analytica (Risso, 2018).

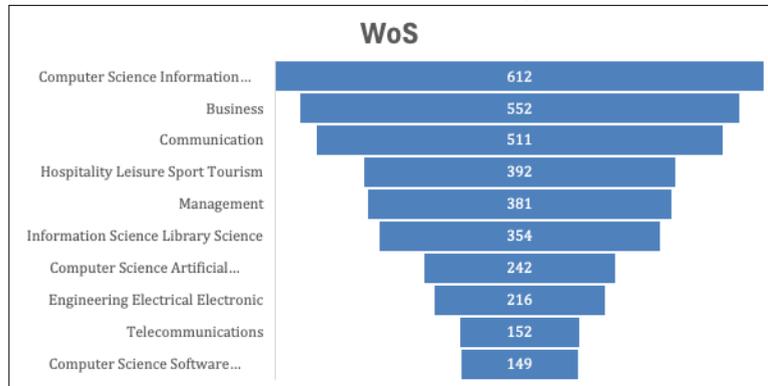
El Gráfico 1 ilustra las dimensiones de *agregación* y *visibilidad* del contenido generado inicialmente por los usuarios y cómo este puede desplazarse dentro de dichas dimensiones a medida que las plataformas lo utilizan, reutilizan, resignifican y vuelven a utilizar, con o sin el consentimiento y/o la conciencia del usuario que lo generó originalmente, en relación con los resultados de tales acciones o los nuevos significados adquiridos por su contenido. Cuando el contenido permanece visible y su significado se conserva en cierta medida, afirmamos que se mantiene dentro del alcance de CGU (parte superior del Gráfico 1), mientras que se trata de DGU cuando se apunta a los usos opacos en los que se atribuyen otros significados, alejados de la mirada y las capacidades cognitivas del usuario (parte inferior del Gráfico 1) y desatando serios cuestionamientos éticos, fundamentalmente en torno a la expectativa de privacidad de los usuarios (Zimmer, 2018).

4. DOS CAMPOS DE INVESTIGACIÓN

Los CGU se han utilizado de forma consistente en algunos campos, como el turismo y los estudios de *marketing*, la comunicación y las ciencias sociales en general. Sin embargo, también son notablemente populares en las ciencias de la información y la computación y en la ingeniería (véanse los gráficos 2 y 3 para las publicaciones históricas de WoS y Scopus). Sin embargo, el enfoque sobre esos campos difiere bastante entre sí. Mientras que en las ciencias sociales hay muchas referencias al concepto central de CGU, como hemos definido anteriormente, en las ciencias políticas, la comunicación, la psicología, etc., en el *marketing* y las ciencias computacionales existe un predominio de los análisis de datos agregados que conducen a nuevos significados, no deseados ni previstos por los usuarios.

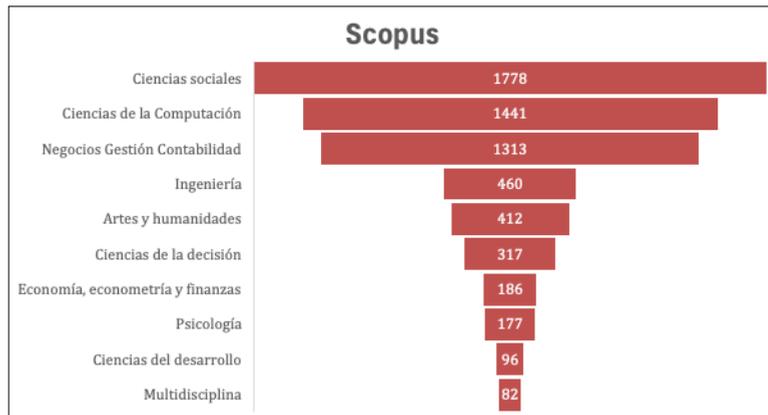
Realizamos un análisis de las 100 primeras citas del artículo que origina esta discusión, aportando una definición unificadora de CGU (Santos, 2022). Los resultados evidencian que el concepto ha sido aplicado en contextos bastante diferentes. Utilizando la distinción actual entre CGU y DGU, observamos que uno de cada tres artículos que mencionan la definición original de CGU se refiere, en realidad, a DGU, es decir, contenido cuyo significado se crea mediante la descontextualización de los insumos originales de los usuarios, es decir, *datos* y no *contenido*.

Gráfico 2. Artículos publicados en revistas indexadas en WoS por categorías temáticas



Fuente: Elaboración de los autores con datos de la búsqueda de Web of Science Core Collection. *Query* original en inglés = (UGC OR “user-generated content” OR “user generated content”); solo artículos; filtrando categorías relacionadas con biomedicina y astronomía, para las cuales el acrónimo arroja otro significado.

Gráfico 3: Artículos publicados en revistas indexadas en Scopus por área temática



Fuente: Elaboración de los autores con datos de la búsqueda de artículos de Scopus. *Query* original en inglés = (UGC OR “user-generated content” OR “user generated content”); solo artículos; filtrando categorías relacionadas con biomedicina y astronomía, para las cuales el acrónimo arroja otro significado.

La inteligencia artificial (IA) es un ejemplo que invita a la reflexión y que conviene analizar en profundidad. Estamos vivenciando últimamente la creación automatizada de grandes cantidades de *contenido generado por inteligencia artificial* (AIGC, por su sigla en inglés), basándose no solo en instrucciones humanas, sino también en contenido disponible en línea, como es el caso de ChatGPT y DALL-E (Cao et al., 2023). Esto añade complejidad

a los métodos de creación de contenido más tradicionales, como el contenido generado por el usuario y el *contenido generado por profesionales* (PGC, por su sigla en inglés), como los trabajadores de *marketing*, por ejemplo. Por lo tanto, plantea desafíos y oportunidades en torno a la seguridad y la privacidad, como la fuga y el seguimiento de datos, los ataques a los modelos y la inseguridad del contenido generado. Pero también abre interrogantes respecto del significado, la propiedad, la autoría, la visibilidad, la interpretación, entre otros.

Ante este escenario, en los últimos años han surgido artículos, leyes y normativas vinculadas al uso de AIGC, notablemente en la Unión Europea y China. Estas medidas en torno a AIGC tienen como objetivo proteger los datos personales en la red, por lo que se ha propuesto regular su uso, salvaguardar la seguridad de los datos y promover el desarrollo prudente de la IA (Chen et al., 2023; Wang et al., 2023). Aunque la AIGC aún no está protegida contra la manipulación y el plagio, se han desarrollado protocolos de protección de derechos de autor y propiedad, considerando que la AIGC se convertirá en la principal forma de creación de contenido en el futuro (Liu et al., 2023; Wu et al., 2023). De ahí la preocupación por problemas relacionados con la IA y la seguridad de los datos y la privacidad (Cao et al., 2023), aunque al mismo tiempo esto no descarta las oportunidades que la IA podría otorgar en campos como las finanzas, la salud, la educación y la industria (Cao, et al., 2025).

¿Cómo aprenden estos sistemas de IA? ¿Cuál es el “contenido” del que los programadores se apropian para enseñar a sus algoritmos? Es posible que este mismo texto, de estar disponible sin muro de pago, fuera capturado por sistemas de IA de forma inadvertida para entrenar uno o muchos algoritmos en el ámbito académico o, en cualquier otro caso, sin consentimiento, sin conciencia o incluso sin intención. Como resultado, podría alimentar el texto de otro autor o autora sin resguardar la referencia original. En una sociedad cada vez más datificada, el contenido generado por los usuarios, independientemente de que se trate de un “me gusta” o de un texto complejo como un artículo científico, puede ser sacado de contexto para construir otros significados, alimentar otros sistemas, generar una existencia paralela fuera de las manos del usuario original.

5. CONSIDERACIONES FINALES

La problemática general de la apropiación de datos en las plataformas digitales conlleva, por lo tanto, no solo las necesidades de avanzar en formas de conceptualizar y operacionalizar las diferentes estrategias y patrones de creación y circulación de contenidos, sino que también plantea cuestionamientos profundos sobre la validez ética de los procedimientos subyacentes. No se puede equiparar la creación consciente y autónoma de los contenidos

con la recolección de datos y metadatos asociados, ni con la resignificación de contenidos, todo ello realizado sin el consentimiento de los individuos que los generaron.

Es cierto que los CGU y los DGU son dos caras de una misma moneda. Sin embargo, una de ellas es brillante, llamativa y visible, mientras que la otra es opaca y está imbricada en áreas grises de regulación de datos y violaciones a la privacidad. Ambas pueden converger en ocasiones, y quizás deberían hacerlo. Pero la distinción conceptual entre ellas resulta fundamental para comprender y garantizar la existencia de recursos dirigidos a ambas dimensiones del contenido generado por los usuarios, separando lo intencional de lo no intencional, lo consciente de lo inconsciente, lo visible de lo opaco. Esto no solo debería potenciar el poder analítico de ambas perspectivas, sino también enfocar la atención hacia un campo más crítico de investigación de Datos Generados por los Usuarios, un campo que dispara preocupaciones éticas sobre la privacidad, la seguridad y la autoría de los datos extraídos de forma subterránea. El presente trabajo pretende contribuir en ese sentido e impulsar dicha agenda bajo una conceptualización detallada del DGU, al delimitarlo propiamente como un objeto de estudio.

REFERENCIAS

- Allan, S. (2007). Citizen Journalism and the Rise of 'Mass Self-Communication': Reporting the London Bombings. *Global Media Journal - Australian Edition*, 1(1), 1-20.
- Bruns, A. (2019). After the 'APicalypse': social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11), 1544-1566. <https://doi.org/10.1080/1369118X.2019.1637447>
- Cadwalladr C. & Graham-Harrison E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P.S. & Sun, L. (2023). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. *ArXiv*, 37(4), 1-44. <https://arxiv.org/pdf/2303.04226>
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. & Sun, L. (2025). A Survey of AI-Generated content (AIGC). *ACM Computing Surveys*, 57(5), 1-38. <https://doi.org/10.1145/3704262>
- Castells, M. (2009). *Comunicación y Poder*. Alianza Editorial.
- Chadwick, A. & Howard, P. N. (2009). Introduction: New directions in Internet politics

- research. In A. Chadwick & P. N. Howard. (eds.), *Routledge handbook of internet politics* (1-12). Routledge.
- Charitsis, V. & Laamanen, M. (2022). When digital capitalism takes (on) the neighbourhood: data activism meets place-based collective action. *Social Movement Studies*, 23(3), 320-337. <https://doi.org/10.1080/14742837.2022.2123314>
- Chen, C., Wu, Z., Lai, Y., Ou, W., Liao, T. & Zheng, Z. (2023). Challenges and remedies to privacy and security in AIGC: Exploring the potential of privacy computing, blockchain, and beyond. *ArXiv*, 37(4),1-43. <https://arxiv.org/pdf/2306.00419>
- Chouliaraki, L. (2010). Ordinary witnessing in post-television news: Towards a new moral imagination. *Critical discourse studies*, 7(4), 305-319.
- Gillmor, D. (2010). *Mediactive*. Dan Gillmor.
- Herrada, N., Santos, M. & Barbosa, S. (2024). Affordances-driven ethics for research on mobile instant messaging: Notes from the Global South. *Mobile Media & Communication*, 12(3). <https://doi.org/10.1177/20501579241247994>
- Jia, L. & Ruan, L. (2020). Going global: Comparing Chinese mobile applications' data and user privacy governance at home and abroad. *Internet Policy Review*, 9(3). <https://doi.org/10.14763/2020.3.1502>
- Johns, A. (2020). 'This will be the WhatsApp election': Crypto-publics and digital citizenship in Malaysia's GE14 election. *First Monday*, 25(12). <https://doi.org/10.5210/fm.v25i12.10381>
- Johns, A., Matamoros-Fernández, A. & Baulch, E. (2024). *WhatsApp: From a One-to-one Messaging App to a Global Communication Platform*. John Wiley & Sons.
- Kaplan, A. M. & Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59-68.
- Kargar, S. & McManamen, K. (2018). Censorship and collateral damage: analyzing the Telegram ban in Iran. *Berkman Klein Center*. <https://cyber.harvard.edu/publication/2018/censorship-and-collateral-damage>
- Liu, Y., Du, H., Niyato, D., Kang, J., Xiong, Z., Miao, C., Shen, X. & Jamalipour, A. (2023). Blockchain-Empowered Lifecycle Management for AI-Generated Content (AIGC) Products in Edge Networks. *ArXiv*, 6(1), 1-9. <https://arxiv.org/pdf/2303.02836>
- Milan, S. & Treré, E. (2021). Big Data From the South(s): An Analytical Matrix to Investigate Data at the Margins. In D. Rohlinger & S. Sobieraj (Eds.), *The Oxford Handbook of Sociology and Digital Media* (pp. 1-21). Oxford University Press.
- Olson, P. (2014). Facebook closes \$19 billion WhatsApp deal. *Forbes*. <https://www.forbes.com/sites/parmyolson/2014/10/06/facebook-closes-19-billion-whatsapp-deal/>

- Pellegrino, G., Söderberg, J. & Milan, S. (2019). Datafication from Below: Epistemology, Ambivalences, Challenges. *Tecnoscienza: Italian Journal of Science & Technology Studies*, 10(1), 89-114. <https://doi.org/10.6092/issn.2038-3460/17431>
- Peng, S. (2024). Datafication and Data Flows. In *International Economic Law in the Era of Datafication* (pp. 195-243). Cambridge University Press.
- Reading, A. (2009). Mobile witnessing: ethics and the camera phone in the 'war on terror', *Globalizations*, 6(1), 61-76. <https://doi.org/10.1080/14747730802692435>
- Rieder, B., Abdulla, R., Poell, T., Woltering, R. & Zack L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring "We are all Khaled Said". *Big Data & Society*, 2(2). <https://doi-org.pucdechile.idm.oclc.org/10.1177/2053951715614980>
- Risso, L. (2018). Harvesting Your Soul? Cambridge Analytica and Brexit. In Jansohn, C. (ed.), *The Selected Proceedings of the Symposium Brexit means Brexit?* (pp. 75-90). Akademie der Wissenschaften und der Literatur, Mainz.
- Ryan, F., Fritz, A. & Impiombato, D. (2020). Front Matter. In TikTok and WeChat: Curating and controlling global information flows (p. [i]-01). *Australian Strategic Policy Institute*. <http://www.jstor.org/stable/resrep26120.1>
- Santos, M. (2022) The 'so-called' UGC: An updated definition of User-Generated Content in the age of Social Media. *Online Information Review*, 46(1), 95-113. <https://doi.org/10.1108/OIR-06-2020-0258>
- Santos, M. (2023). Politicizing witnessing: Testimonial user-generated content in the aftermath of Rousseff's impeachment in Brazil. *Convergence: The International Journal of Research into New Media Technologies*, 29(6), 1517-1534. <https://doi.org/10.1177/13548565231164760>
- Santos, M., Saldaña, M. & Tsyganova, K. (2024). Subversive Affordances as a Form of Digital Transnational Activism: The case of Telegram's Native Proxy. *New Media & Society*, 26(1). <https://doi.org/10.1177/14614448211054830>
- Saura, J. R., Ribeiro-Soriano, D. & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, 60, 1-13. <https://doi.org/10.1016/j.ijinfomgt.2021.102331>
- Saura, J. R., Palacios-Marqués, D. & Ribeiro-Soriano, D. (2023). Privacy concerns in social media UGC communities: Understanding user behavior sentiments in complex networks. *Information Systems and e-Business Management*, 23, 125-145. <https://doi.org/10.1007/s10257-023-00631-5>
- Segura, M. S. & Waisbord, S. (2019). Between Data Capitalism and Data Citizenship. *Television & New Media*, 20(4), 412-419. <https://doi-org.pucdechile.idm.oclc.org/10.1177/1527476419834519>

- Setzer, V. W. (2004). Dado, informação, conhecimento e competência. *Folha Educação*, 27, 6-7. <https://www.ime.usp.br/~vwsetzer/dado-info-Folha.html>
- Srnicek, N. (2017). *Platform capitalism*. Polity Press.
- Quandt, T. (2018). Dark Participation. *Media and Communication*, 6(4), 36-48. <https://doi.org/10.17645/mac.v6i4.1519>
- Treré, E. (2020). The banality of WhatsApp: On the everyday politics of backstage activism in Mexico and Spain. *First Monday*, 25(12). <https://doi.org/10.5210/fm.v25i12.10404>
- van der Vlist, F. N. & Helmond, A. (2021). How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society*, 8(1). <https://doi.org/10.1177/20539517211025061>
- van Dijck, J. (2009). Users like you? Theorizing agency in user-generated content. *Media, Culture, and Society*, 31(1), 41.
- van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.
- Wang, T., Zhang, Y., Qi, S., Zhao, R., Xia, Z. & Weng, J. (2023). Security and privacy on generative data in AIGC: A survey. *ACM Comput*, 1(1). <https://arxiv.org/pdf/2309.09435>
- Wardle, C., Dubberley, S. & Brown, P. (2014). Amateur footage: a global study of user-generated content in TV and online news output. *Tow Center for Digital Journalism. A Tow/Knight Report*, <http://usergeneratednews.towcenter.org/wp-content/uploads/2014/05/Tow-Center-UGC-Report.pdf>
- Wardle, C. & Derakhshan, H. (2017). Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. *Council of Europe*. <https://rm.coe.int/information-disorder-report-november-2017/1680764666>
- Williams, A., Wardle, C. & Wahl-Jorgensen, K. (2011). "Have they got news for us?" Audience revolution or business as usual at the BBC? *Journalism Practice*, 5(1), 85-99.
- Wu, J., Gan, W., Chen, Z., Wan, S. & Lin, H. (2023). Ai-generated content (AIGC): A survey. *arXiv preprint*. <https://arxiv.org/pdf/2304.06632>
- Wunsch-Vincent, S. & Vickery, G. (2007). Participative Web and User-Created Content: Web 2.0, Wikis and Social Networking. *OECD*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2007/09/participative-web-and-user-created-content_g1gh826c/9789264037472-en.pdf
- Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media + Society*, 4(2). <https://doi.org/10.1177/2056305118768300>

* Contribución de autoría: la conceptualización y el desarrollo inicial del primer autor. La segunda autora realizó análisis de datos y colaboró en la revisión de literatura pertinente y la redacción del artículo

* Nota: el Comité Académico de la revista aprobó la publicación del artículo.

* El conjunto de datos que apoya los resultados de este estudio no se encuentran disponibles para su uso público. Los datos de la investigación estarán disponibles para los revisores, si así lo requieren.



Artículo publicado en acceso abierto bajo la Licencia Creative Commons - Attribution 4.0 International (CC BY 4.0).

IDENTIFICACIÓN DE LOS AUTORES

Marcelo Santos. Doctor en Ciencias de la Comunicación por la Pontificia Universidad Católica de Chile (Chile). Profesor asociado de la Facultad de Comunicación y Letras, Universidad Diego Portales (Chile). Investigador del Centro de Investigación en Comunicación, Literatura y Observación Social y del Núcleo Milenio para el Estudio de la Política, Opinión Pública y los Medios en Chile, Universidad Diego Portales. Investigador del Instituto Nacional de Ciência e Tecnologia para la Democracia Digital (Brasil). Ha publicado artículos en revistas como *New Media & Society*, *Political Communication, Information, Communication & Society*, *Social Media & Society*, *Mobile Media & Communication*, *Digital Journalism*, entre otras.

Luján Román Aponte. Doctora –candidata– en Ciencias de la Comunicación, Pontificia Universidad Católica de Chile (Chile). Docente en la Universidad Finis Terrae (Chile), en la Universidad de Chile (Chile) y en la Universidad Gabriela Mistral (Chile).