

# User-generated content or data?

The (in)voluntary dynamics of online content generation<sup>1</sup>

¿Contenido o datos generados por el usuario?

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

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>

## ► LUJÁN ROMÁN APONTE

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

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

**HOW TO CITE:** 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>

Submission date: December 7, 2024

Acceptance date: May 6, 2025

<sup>1</sup> The first author, Marcelo Santos, received funding from the National Agency for Research and Development (ANID) through the National Fund for Scientific and Technological Development, No. 11230980, and the Funding Fund for Research Centers in Priority Areas, No. NCS2024\_007. The second author, Luján Román Aponte, received funding from the National Agency for Research and Development (ANID) through the National Doctoral Scholarship, 2020, No. 21200327.

## 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:** *datification, social media, user-generated content, user-generated data, privacy.*

## RESUMEN

La creciente datificación 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:** *datificación, redes sociales, Contenido Generado por Usuario, datos generados por el usuario, privacidad.*

## 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:** *datificação, redes sociais, conteúdo gerado por usuário, dados gerados pelo usuário, privacidade.*

## 1. INTRODUCTION

The emergence of digital technologies that allow, on the one hand, the creation of content in various formats (photos, texts, etc.), and which, in turn, are articulated with other components that enable exchange through their own, autonomous channels (Castells, 2009; Santos, 2022) has precipitated an unprecedented production of independent content in the history of media. However, almost all of what is published is, literally, “contained” on private commercial platforms that are, to some extent, for-profit. This factor is relevant and has significant effects on data appropriation processes beyond the content originally published or the user’s intention when publishing. We therefore hypothesize that a conceptual gap is emerging in the field, that encompasses content and meanings extracted without consent, awareness, or intention on the part of users on such platforms. To address this gap, we propose the conceptualization and operationalization of the term User-Generated Data (UGD) as a necessary concept to support an entire line of critical research on digital platforms, within a media ecosystem that far exceeds the content consciously generated by its authors.

The year 2022 marked the 20th anniversary of the introduction of the term User-Generated Content (UGC) into the vocabulary of the media and academia. With the aim of establishing an adequate definition that would allow for the articulation of different fields within the scope of this concept, the notion proposed at the time was:

User-Generated Content is any is any kind of text, data or action performed by online digital systems users, published and disseminated by the same user through independent channels, that incur an expressive or communicative effect either on an individual manner or combined with other contributions from the same or other sources. (Santos, 2022, p. 108)

Over the past few decades, UGC has been at the heart of communicative practices by ordinary (Chouliaraki, 2010) non-professional users, such as citizen journalism and the witnessing of extraordinary events (Santos, 2023), among others. Behind the popularization of smartphones are economic, social, and cultural factors (Wunsch-Vincent & Vickery, 2007), but also technological advances that have enabled these changes: first, multimedia functions were incorporated, allowing recording, and then, devices became truly smart and almost permanently connected to telematic networks. Starting with pixelated amateur recordings of extraordinary events such as the 2005 London bombings (Allan, 2007; Reading, 2009) and the 2006 Thailand tsunami (Wardle, Dubberley & Brown, 2014), phone-generated content gradually became a real alternative for journalists, especially when the location being reported on was remote or dangerous (Wardle & Derakhshan, 2017). The production

and distribution of user-generated content grew exponentially, and managing citizen contributions as news sources became an increasingly pressing issue for newsrooms. This led the British Broadcasting Corporation (BBC), for example, to implement what they called the “BBC UGC Hub” (Williams, Wardle & Wahl-Jorgensen, 2011) in 2005 to cope with the volume of citizen content directed at them through digital channels; not only to separate the wheat from the chaff, but to verify the reliability of the selected content.

In the early 2000s, many analysts’ perceptions of the autonomous and distributed creation of content by ordinary citizens were optimistic. For example, it was argued that citizens with smartphones would supposedly become *little brothers* (Chadwick & Howard, 2009), alluding to the figure of “Big Brother,” or that they would build a mosaic of fragments of user-generated content that could potentially become an inverted (bottom-up) panopticon, in which social media would become a weapon of citizenship and a guardian of institutionality, including traditional media and police violence (Santos, 2023).

The possibility of autonomously creating and publishing content (Castells, 2009) was, to some extent, associated with the normative view that exercising citizenship was equivalent to participating in these flourishing new digital environments as *media-active* citizens (Gillmor, 2010). This perception, this kind of techno-optimistic vision, ignored that the vast majority of users do not create content, they only use, in various ways, content created by others (van Dijck, 2009). Later, this perspective was contrasted with evidence on the low effectiveness of the UGC as an alternative narrative in the face of extraordinary events, such as street protests, demonstrated by Santos (2023) in his analysis of Twitter (nowadays X). Finally, there was a radical shift, from this initial perception of civic purity/positivity in the multiple forms of citizen participation through digital platforms to the acknowledgement of various forms of “dark participation” that took over a good part of the interactions on social networks, as Quandt (2018) operationalized.

Throughout the 21st century, the increasing datafication of life has brought with it new ways of assigning meaning to social problems, serving as a tool to drive more efficient policymaking. In other words, various dimensions of life, or their decision-making processes, are managed by data generated in the context of citizens’ digital activity. Milan and Treré (2021) highlight the risks associated with increasing social disparity and citizen exploitation, stemming from the intrinsic paradoxes of the lack of data on vulnerable populations (Peng, 2024). In this context, we suggest a deeper analysis not only of content creation but also of the production of meaning that occurs on, or mediated by, digital platforms. And not only regarding to the creation of content by the users who create these forms of content, but also to the unexpected, even unauthorized, meanings produced by third parties when they appropriate such content or its associated metadata. Different forms of data collection, extraction,

visualization, and analysis, not without criticism, have emerged or become more prominent (Milan & Treré, 2021; Pellegrino, Söderberg & Milan, 2019; Rieder et al., 2015), as our sociability became a valuable commodity in the 21st century (van Dijck, 2013).

We academics are not exempt from such criticism, as we are part of the system dedicated to framing the problems of our time. Not only have many of us been part of the techno-optimist narrative, but scientific enthusiasm and curiosity may have driven *data extractivist* practices with an ethical sense that is inattentive to the specificities of the digital world (Herrada, Santos & Barbosa, 2024). However, academic work also has the responsibility to continue working on these problems, even if it means contradicting reflections formulated in the past.

We will demonstrate in this article that the term UGC has encapsulated two different fields. As we advance to a more detailed distinction between the two, it will become clear that there are lines of research that resonate in each of them. This article is, therefore, a critical reflection on the literature, but also on our own previous conceptualization of UGC (Santos, 2022). The result is its division into a dual concept: User-Generated Content and User-Generated Data, two objects of study and two fields.

This article challenges the idea behind the “content” part of the acronym and problematizes it in that context, asserting that there are factors that align better with the concept of “data.” To arrive at this distinction, we will first conduct a brief archaeology of UGC and how it became a commodity; then we will discuss the distinction between content and data, and finally we will delineate the fields for UGC and UGD.

## 2. COMMODIFICATION OF UGC

Although the literature in some disciplines has neglected the operationalization of social media content as user-generated content (UGC), we have previously stated that UGC is the sociotechnical element at the epicenter of social media (Santos, 2022). Social networks such as X, Facebook, Instagram, or TikTok gain meaning due to the contributions not only of prominent actors such as celebrities, institutions, etc., but also of ordinary users. While the former are often professionalized<sup>2</sup>, the latter feed these platforms with content, motivated by some degree of communicative purpose, such as increased visibility, networking, among others. In this sense, UGC resembles the functioning of the *fandom* world, in terms of the generation and distribution of content generated by “people who gather in groups based on common preferences and interests” (Torti & Schandor, 2013, p. 3), which are leveraged by the entertainment industry to obtain greater economic profit. This process

<sup>2</sup> Highly visible users use communication and marketing professionals and professional community managers.

is similar to what has occurred in recent decades with user data, transformed into a commodity, as sociability has become a product (van Dijck, 2013; Srnicek, 2017) in an economic context in which much attention has been paid to data extraction.

This widespread practice of data extraction by social media platforms, political agencies, marketing agencies, among others, and its progressive normalization (Segura & Waisbord, 2019) has been commonly identified as part of the *datafication of society* (Milán & Treré, 2021), framed within an economy characterized as *platform capitalism* (Srnicek, 2017). In this datafied economy, only a few platforms hold the market due to, among other business practices, the network effect, a phenomenon that allows platforms to become more valuable thanks to the massive interaction of users who utilize the tools and functionalities provided by digital environments. This effect has very little to do with the widely defended values of the free market: the more activities users carry out on the platforms, the more these companies reinforce their monopoly by centralizing data within their environment (Srnicek, 2017). As an example of this, when exploring the businesses, partnerships, and partner integrations of the 20 most used social networks, van der Vlist and Helmond (2021) noted the existence of an integrated platform ecosystem rather than a single platform, where governance and control are exercised through partnership and infrastructure agreements with data intermediaries that map individuals' digital footprints, deepening the process of platform capitalization. To make matters worse, these extremely data-rich environments are often repurposed for aims that were not originally declared, entering into areas of gray regulation and objectionable ethics.

Meta companies are perhaps the clearest example. The Cambridge Analytica scandal in 2016 (Cadwalladr & Graham-Harrison, 2018) was a flagrant display of users' lack of control over their privacy through a questionable practice of psychological profiling and micro-targeting of political propaganda through Facebook (Risso, 2018). At the same time, it served as a pretext for the company to close its API, further limiting the ability of independent observers such as journalists or academics to monitor, study, and scrutinize its activity (Bruns, 2019). A few years later, Meta acquired WhatsApp for the extraordinary sum of \$19 billion (Olson, 2014), with the promise to maintain some strategic features of the app, for example, its ad-free model and its privacy standards. However, shortly after these promises, Meta began investing in the WhatsApp platform, compromising user privacy in many ways, from sharing their meta-data with Facebook to enhance its commercial appeal to allowing companies to collect user data through the WhatsApp business environment (Johns, Matamoros-Fernández & Baulch, 2024). Data activism campaigns such as #blocksidewalk and *Fuck Off Google*, driven by activists resisting the commodification of their data resulting from its appropriation by powerful tech corporations

through unethical practices (Charitsis & Laamanen, 2024), underscore the need to look beyond the mere functionality of digital technologies as tools for organizing social movements—such as in the case of the Arab Spring—and to explore alternative pathways within the same data capitalism framework. Supposedly protected environments of encrypted applications may create conditions for safe activism such as the “backstage activism” proposed by Tréré (2020) or the articulation of dissent in authoritarian contexts such as Malaysia (Johns, 2020) or Russia (Santos, Saldaña & Tsyganova, 2024). However, users’ perceptions of security and privacy often do not align with the actual protections offered by these platforms (Herrada, Santos, & Barbosa, 2024). Therefore, it is suggested that users adapt their behavior in digital environments according to their perceptions of trust and privacy in each digital ecosystem, in order to feel more secure (Saura, Palacios-Marqués & Ribeiro-Soriano, 2023).

In this world of Big Data, multiple data collection and analysis strategies are emerging, while society as a whole tries to catch up in normative, regulatory, and even ethical terms. As the race progresses, many gray areas are emerging, where private platforms have been collecting vast amounts of data to turn it into products, while state platforms collect it not only to develop public policies but also to survey and exercise control over their citizens. Examples of the latter include the push for Iran’s state-run messaging platform Soroush to replace Telegram (Kargar & McManamen, 2018), or popular apps in China such as Douyin, the local variant of TikTok, and WeiXin, the domestic version of WeChat, which vary in privacy and data protection standards across both domestic and international versions (Jia & Ruan, 2020). Chinese companies running the apps are required to adhere to these standards regarding the content they can publish, so they invest in content filtering and human curation systems (Ryan, Fritz, & Impiombato, 2020).

As we move from intentional, even naive, forms of user content production that ultimately seek to promote some form of sociability, to opaque procedures for collecting reinterpreted data with little or no awareness on the part of those who generate it, we observe a shift from user-generated *content* to user-generated *data*. From the perspective of the user, it will be argued in the following section that “data” refers to meaning constructed by third parties, often at the expense of the users who generate it, while “content” refers to data created and inexorably linked to context, which gives it a meaning or purpose, to some extent, intended by the users who generate it.

### 3. FROM UGC TO UGD

While some content easily fits into the UGC category, such as images or texts created and published by users through their own channels (Santos, 2022), such as individual Twitter/X accounts (Santos, 2023), other types of

information do not clearly reflect users' intention to share or create content. The main dimension of this discussion lies in the tension between, on the one hand, content creation that entails a certain awareness and intention on the part of the user, and, on the other, information production that relies on the recontextualization, aggregation, and redefinition of UGC, giving rise to new content aimed at audiences other than those originally intended. In other words, what content are users *generating*, and what content is *being generated* behind their backs?

The few definitions of UGD present in the literature are usually conditioned by the study's discipline of origin, such as the definition by Saura, Ribeiro-Soriano, and Palacios-Marqués (2021) in the field of marketing and innovation: "UGD includes all forms of information and data that users individually generate as a result of interacting with the elements that make up any digital marketplace (actions, experiences, feelings, comments, reviews, etc.)" (p. 1). The authors then set out to discuss the limits of privacy in the appropriation of UGD in innovation-driven initiatives, but their definition is not adequately distinguished from common definitions of UGC, limiting it to "digital marketplaces." An example of a definition of UGC that is not clearly distinguishable from Saura, Ribeiro-Soriano, and Palacios-Marqués's definition of UGD is Kaplan and Haenlein's (2010) definition: "User-generated content (UGC) can be viewed as the sum of all the ways in which people use social media" (p. 61). Another frequently cited definition of user-generated content refers to: "1) content that is made publicly available on the internet, 2) that reflects a certain amount of creative effort, and 3) that is created outside of professional routines and practices" (Wunsch-Vincent & Vickery, 2007, p. 9). In this definition, user-generated data is merely a loosely defined subset of user-generated content.

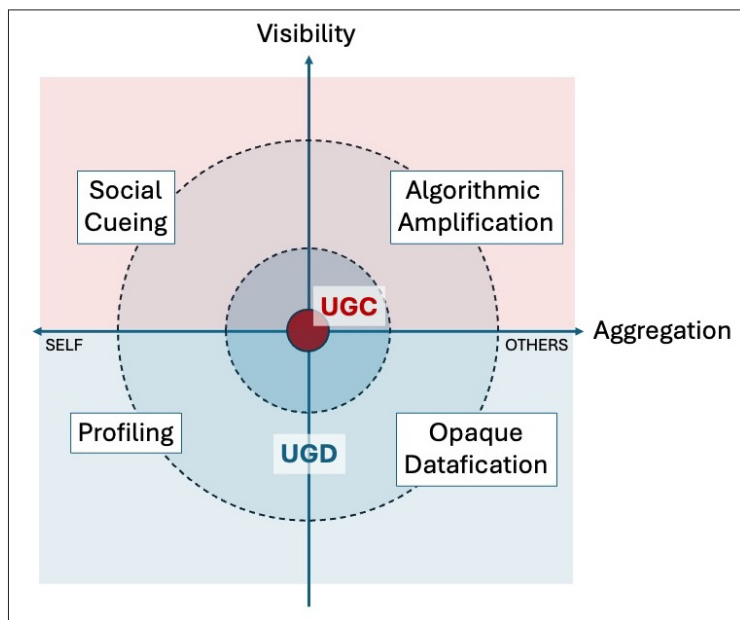
But how can we distinguish *data* from *content*? In information science, a key distinction is made between data and information: while the latter is defined as contextualized data, the former represents the raw form of any digital entry recorded in a database (Setzer, 2004). In this framework, *information* is closely related to what we consider content, while data does not necessarily retain its link to the original context or may refer to a context different from the original without (or with very little) awareness or intention on the part of the user(s) who generate it.

As such, user-generated data refers to data generated from activities on digital platforms that acquire meaning as they are processed differently than originally intended (extracted, recontextualized, aggregated, etc.), used to generate new meanings, with little or no awareness on the part of users. Thus, while UGC is originally generated by ordinary users, published on their own individual channels, and its trajectory is visible to its author, UGD disarticulates the original meaning of the content, violating principles of privacy, intent,



and awareness. Therefore, the data in UGD is transformed into information to the extent that it is recontextualized and repurposed by data extractors. In other words, while UGC implies awareness of the visibility and control over the context of the generated content, UGD ignores both dimensions, leading to different scenarios, represented in Figure 1. At the center is the original UGC. As external actors—such as other users, algorithms, marketing agencies, data extractors, etc.—act upon it, the content can fluctuate across the different dimensions, as illustrated in the following diagram.

**Figure 1.** Map of possibilities that distinguish user-generated content (UGC) from user-generated data (UGD) along two dimensions: visibility and aggregation. While visible content that retains its original meaning is considered UGC, content that is repurposed through opaque means of processing is considered UGD



**Source:** Created by the authors.

When user-generated content goes viral “organically” with the help of algorithms and other user data (e.g., “likes,” “shares,” etc.), it largely retains its original meaning, and the user can observe how it circulates even if that visibility is amplified by external factors. This exposure can have both positive and negative consequences, and while they are not always necessarily negative consequences, they are certainly not deliberately intended. As content accumulates “algorithmic amplification,” the user loses control over the content. Another type of aggregation of what was originally UGC is the cumulative

collection of data as sources for large-scale data mining, in ways that are invisible to the user. In these cases, the intended meaning of the content is lost, and its use, analysis, and interpretation become opaque to the majority of society. On the other hand, when aggregation occurs at the user level, the visible face of the aggregated data is something like a social media resume, that is, the social signals that inform user perceptions within a community. The opaque version of such aggregation is the creation of invisible (and therefore unconscious) profiles of users, as exemplified by the Cambridge Analytica scandal (Risso, 2018).

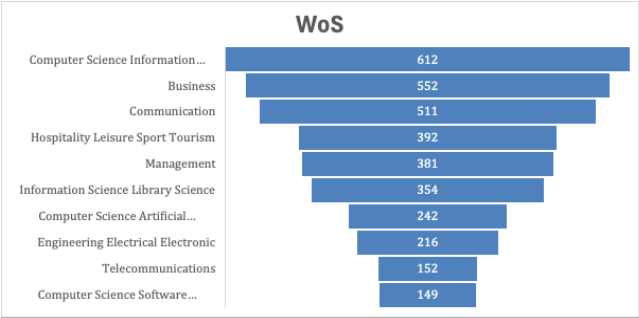
Figure 1 illustrates the dimensions of *aggregation* and *visibility* of content generated initially by users, and how it can shift within these dimensions as platforms use, repurpose, resignify, and reuse it, with or without the consent and/or awareness of the user who generated it originally, in relation to the outcomes of such actions or the new meanings acquired by its content. When content remains visible and its meaning is retained to some extent, we argue that it falls within the scope of UGC (top of Figure 1), while it is considered UGD when it refers to opaque uses in which other meanings are attributed, removed from the user's gaze and cognitive capacities (bottom of Figure 1), and triggering serious ethical questions, mainly around users' expectation of privacy (Zimmer, 2018).

#### 4. TWO FIELDS OF RESEARCH

UGCs has been consistently used in some fields, such as tourism and marketing studies, communication, and social sciences in general. However, it is also notably popular in information and computer sciences and engineering (see Figures 2 and 3 for historical WoS and Scopus publications). However, the approach in these fields differs considerably. While in the social sciences there are many references to the central concept of UGC, as defined above, in political science, communication, psychology, etc., in marketing and computer science, there is a predominance of aggregated data analyses that lead to new meanings, neither intended nor anticipated by users.

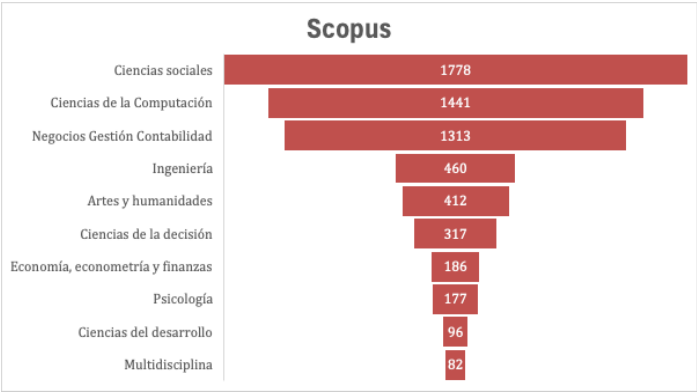
We conducted an analysis of the first 100 citations of the article that originated this discussion, providing a unifying definition of UGC (Santos, 2022). The results show that the concept has been applied in quite different contexts. Using the current distinction between UGC and UGD, we observe that one in three articles that mention the original definition of UGC actually refers to UGD, that is, content whose meaning is created by decontextualizing the original user input, i.e., *data* rather than *content*.

Figure 2. Articles published in journals indexed in WoS by subject category



**Source:** Created by the authors using data from the Web of Science Core Collection search. Original query in English = (UGC OR “user-generated content” OR “user generated content”); only articles; filtering categories related to biomedicine and astronomy, for which the acronym holds another meaning

Figure 3: Articles published in journals indexed in Scopus by subject area



**Source:** Created by the authors using data from the Scopus article search. Original query in English = (UGC OR “user-generated content” OR “user generated content”); only articles; filtering categories related to biomedicine and astronomy, for which the acronym holds another meaning.

Artificial intelligence (AI) is a thought-provoking example that deserves in-depth analysis. We are recently experiencing the automated creation of large amounts of *AI-generated content* (AIGC), based not only on human instructions but also on content available online, as is the case with ChatGPT and DALL-E (Cao et al., 2023). This adds complexity to more traditional content creation methods, such as user-generated content and *professionally generated content* (PGC), such as marketing professionals, for example. It therefore poses challenges and opportunities around security and privacy, such as data leakage

and tracking, attacks on models, and the insecurity of the generated content. But it also raises questions about meaning, ownership, authorship, visibility, interpretation, and more.

Given this scenario, in recent years, articles, laws, and regulations related to the use of AIGC have emerged, notably in the European Union and China. These measures surrounding AIGC aim to protect personal data online, and therefore, it has been proposed to regulate its use, safeguard data security, and promote the prudent development of AI (Chen et al., 2023; Wang et al., 2023). Although AIGC is not yet protected against manipulation and plagiarism, copyright and ownership protection protocols have been developed, considering that AIGC will become the primary form of content creation in the future (Liu et al., 2023; Wu et al., 2023). Hence, concerns arise about issues related to AI and data security and privacy (Cao et al., 2023), although at the same time, this does not rule out the opportunities that AI could offer in fields such as finance, healthcare, education, and industry (Cao et al., 2025).

How do these AI systems learn? What is the “content” that programmers appropriate to teach their algorithms? It’s possible that this same text, if available without a paywall, could be inadvertently captured by AI systems to train one or more algorithms in the academic field, or in any other case, without consent, awareness, or even intention. As a result, it could feed into another author’s text without safeguarding the original reference. In an increasingly data-driven society, user-generated content, whether it’s a “like” or a complex text like a scientific paper, can be taken out of context to construct other meanings, feed into other systems, and generate a parallel existence outside the hands of the original user.

## 5. FINAL CONSIDERATIONS

The general issue of data appropriation on digital platforms therefore entails not only the need to advance ways of conceptualizing and operationalizing the different strategies and patterns of content creation and circulation, but also raises profound questions about the ethical validity of the underlying procedures. The conscious and autonomous creation of content cannot be equated with the collection of data and associated metadata, nor with the redefinition of content, all of which is carried out without the consent of the individuals who generated it.

It is true that UGC and UGD represent two sides of the same coin. However, one is bright, eye-catching, and visible, while the other is opaque and embedded in gray areas of data regulation and privacy violations. The two can sometimes converge, and perhaps they should. But the conceptual distinction between them is essential to understanding and ensuring the existence of resources directed at both dimensions of user-generated content, separating the intentional from the unintentional, the conscious from the unconscious, the visible from the opaque. This should not only enhance the analytical power of both perspectives, but also

focus attention on a more critical field of User-Generated Data research—one that raises pressing ethical concerns about the privacy, security, and authorship of data that is extracted covertly. This paper aims to contribute in this regard and advance this agenda through a detailed conceptualization of UGD, by properly delineating it as an object of study.

## REFERENCES

- 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](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>

\* Author contribution: the conceptualization and initial development of the paper was carried out by the first author. The second author performed data analysis and collaborated in the review of relevant literature and the writing of the article.

\* Note: the Academic Committee of the journal approved the publication of the article.

\* The dataset that supports the results of this study is not available for public use. The research data will be made available to reviewers, if required.



Article published in open access under the Creative Commons License - Attribution 4.0 International (CC BY 4.0).

#### AUTHOR IDENTIFICATION

**Marcelo Santos.** PhD in Communication Sciences by the Pontificia Universidad Católica de Chile (Chile). Associate Professor at the Faculty of Communication and Letters, Universidad Diego Portales (Chile). Researcher at the Center for Research in Communication, Literature, and Social Observation and the Millennium Nucleus for the Study of Politics, Public Opinion, and the Media in Chile, Universidad Diego Portales. Researcher at the National Institute of Science and Technology for Digital Democracy (Brazil). He has published articles in journals such as *New Media & Society*, *Political Communication*, *Information, Communication & Society*, *Social Media & Society*, *Mobile Media & Communication*, *Digital Journalism*, among others.

**Luján Román Aponte.** PhD candidate in Communication Sciences, Pontificia Universidad Católica de Chile (Chile). Professor at the Universidad Finis Terrae (Chile), the Universidad de Chile (Chile), and the Universidad Gabriela Mistral (Chile).