

# Contenido basura en internet: Factores, efectos y propuestas de solución

Rafael Guzman-Cabrera\*

Universidad de Guanajuato,  
Campus Irapuato-Salamanca,  
Departamento de Ingeniería Eléctrica,  
Mexico

guzmanc@ugto.mx

**Resumen.** En los últimos años el internet se ha consolidado como el principal medio de comunicación y difusión de información en mucho debido a la sorprendente expansión que ha tenido, lo que simultáneamente ha propiciado una proliferación de contenido basura. Pero, ¿qué es el contenido basura? Pues el contenido al que me refiero abarca información irrelevante, falsa o intencionalmente manipulada, incluyendo la *misinformation*, *disinformation*, *fake news*, periodismo amarillista, y materiales sintéticos generados mediante Inteligencia Artificial. El crecimiento de este fenómeno se atribuye a factores tanto tecnológicos como socioeconómicos. El impacto de esta problemática es muy amplio, ya que va desde la disminución de la confianza en instituciones y medios por parte de los usuarios; intensifica la división social por medio de la polarización entre grupos. también contribuye a la saturación informativa e influye de manera determinante en decisiones cruciales en ámbitos como la política y la salud pública. Para disminuir la cantidad de contenido basura en internet se requiere un enfoque multidimensional que integre el desarrollo de filtros avanzados basados en aprendizaje automático, que existan normas que lo regulen y, fundamentalmente, el fortalecimiento de la alfabetización mediática como mecanismo de resiliencia ciudadana frente a la desinformación.

**Palabra clave.** Contenido basura, desinformación, noticias falsas.

## Junk Content on the Internet: Factors, Effects and Proposed Solutions

**Abstract.** In recent years, the internet has become the primary means of communication and information

dissemination, largely due to its astonishing expansion, which has simultaneously led to a proliferation of junk content. But what exactly is junk content? The content I'm referring to encompasses irrelevant, false, or intentionally manipulated information, including misinformation, disinformation, fake news, sensationalist journalism, and synthetic materials generated by Artificial Intelligence. The growth of this phenomenon is attributed to both technological and socioeconomic factors. The impact of this problem is far-reaching, ranging from decreased user trust in institutions and media; it intensifies social division through polarization between groups; it also contributes to information overload; and it significantly influences crucial decisions in areas such as politics and public health. To reduce the amount of junk content on the internet, a multidimensional approach is required that integrates the development of advanced filters based on machine learning, the existence of regulations, and, fundamentally, the strengthening of media literacy as a mechanism for citizen resilience against disinformation.

**Keywords.** Junk content, misinformation, disinformation, fake news.

## 1. Introducción

Desde hace unos años, el número de personas, cada vez mayor, que utilizamos el internet de manera cotidiana aunado a la cantidad de servicios y dispositivos que requieren de esta conexión para operar que van desde teléfonos móviles y asistentes personales hasta sistemas de posicionamiento global de uso cotidiano para facilitar el desplazamiento de personas en ciudades y carreteras, los cuales ya funcionan



Fig. 1. Clasificación del contenido basura en internet

como red social, han hecho que el internet se convierta en el principal medio de interacción humana y el repositorio primario del conocimiento global.

Miles de millones de individuos acceden diariamente a plataformas digitales con fines informativos, comunicativos y de entretenimiento. De hecho, algunos autores lo han llamado, desde hace unos años, la sociedad del conocimiento, ya que en internet encontramos información de prácticamente cualquier área del conocimiento y en cualquier idioma o dialecto. Aunque aquí juega un papel fundamental el número de usuarios por idioma, más usuarios implica mayor cantidad de información en ese idioma.

Sin embargo, este ecosistema masivo ha propiciado la diseminación incontrolada del contenido basura también conocido como *garbage content*. Este término involucra desde mensajes redundantes o irrelevantes hasta la desinformación política, teorías conspirativas, correo *spam*, y materiales sintéticos de alta fidelidad producidos por la IA generativa sin supervisión ética o control humano (Wardle & Derakhshan, 2017).

Es claro que este volumen de información en formato electrónico, que, además, crece día a día, requiere del desarrollo de nuevas herramientas que permitan su manejo y en el tema que nos ocupa en este trabajo de estrategias que permitan disminuir la proliferación de contenido basura en internet.

Pensémoslo de esta manera: todos aportamos información a la internet día a día y nadie revisa lo que subimos, aunado a la irrupción de generadores avanzados de texto, imágenes y

video basados en IA ha escalado la capacidad de crear y distribuir material digital indistinguible del contenido auténtico, lo cual plantea profundos dilemas éticos y de veracidad (Hao, 2022).

La diseminación del contenido basura es un proceso multifactorial, resultante de la convergencia de dinámicas tecnológicas, económicas y sociales. Dentro de los factores tecnológicos tenemos por un lado la democratización, esto es el hecho de que existen una gran variedad de herramientas que nos permiten realizar la publicación de contenido en internet y por otro lado tenemos la falta de filtros de calidad y la velocidad de difusión de las redes sociales.

Mientras que por el lado de los factores económicos y sociales tenemos por ejemplo a los algoritmos de recomendación, los cuales buscan maximizar el tiempo de permanencia del usuario, priorizando contenido polémico o emocionalmente cargado por encima de su veracidad (Tufekci, 2015).

El impacto del contenido basura se manifiesta en algunos sectores de la sociedad de manera crítica, por ejemplo, en el sector político, la desinformación ha demostrado su capacidad para distorsionar procesos electorales, polarizar el debate y mermar la confianza institucional (Allcott & Gentzkow, 2017; Lazer et al., 2018).

En el ámbito de la salud y la ciencia, la difusión de datos falsos o malinterpretados puede acarrear riesgos para la salud pública, como la resistencia a la vacunación o la promoción de prácticas pseudocientíficas (Pulido et al., 2020). A nivel individual, la exposición constante a ruido informativo genera fatiga informativa, un estado de saturación que desemboca en la desconfianza generalizada y la suspensión del pensamiento crítico (Pennycook & Rand, 2019).

Desde una perspectiva económica, este contenido es el pilar de una industria basada en la economía de la atención. Los modelos de negocio, sustentados en la publicidad y la recolección de datos, incentivan la producción masiva de material superficial diseñado para generar clics (Wu, 2016).

Esto impone a las grandes corporaciones un dilema ético: la optimización del compromiso del usuario amplifica inherentemente la visibilidad del contenido polarizante (Vaidhyanathan, 2018).

**Tabla 1.** Taxonomía del contenido basura por intencionalidad

Tipo	Intencionalidad	Descripción	Ejemplo
<b>Misinformation</b>	No existe una mala intención (el emisor cree que es cierta).	Información falsa, inexacta o engañosa, difundida sin ánimo de causar daño deliberado (Pennycook et al., 2021). Aquí más bien es ignorancia o una cuestión ideología del usuario.	Difusión errónea de una cura casera para una enfermedad crónica.
<b>Disinformation</b>	Intencional (creada para engañar).	Información falsa o inexacta, fabricada y difundida con la intención deliberada de manipular la opinión pública o perpetrar fraudes (Naples, 2019).	Campañas políticas (granja de bots) para alterar narrativas mediáticas.
<b>Fake News</b>	Intencional (fabricada con apariencia noticiosa).	Contenido noticioso fabricado o extremadamente inexacto que imita la estética de los medios legítimos, buscando confundir y engañar (Lazer et al., 2018).	Noticias sobre el origen conspirativo de pandemias o desastres naturales.
<b>Hyperpartisan News</b>	Intencional (sesgo ideológico).	Cobertura de eventos reales con una distorsión o sesgo ideológico extremo, resultando en una forma de <i>misinformation</i> engañosa (Corbu et al 2021).	Medios que solo resaltan los logros de un gobierno, ocultando o minimizando sus fracasos.
<b>Periodismo Amarillista (Yellow Journalism)</b>	Intencional (creado para atraer audiencia/clics).	Contenido sensacionalista, superficialmente investigado, que sacrifica la precisión por el impacto emocional para atraer audiencia (Pennycook et al., 2021).	Titulares escandalosos sobre celebridades o exageración de estudios médicos para vender suplementos.

En respuesta a este panorama desalentador, el estudio del contenido basura se aborda desde la informática, a través de algoritmos de detección automática de desinformación (Shu et al., 2020); desde la educación, mediante la promoción de la alfabetización mediática; y desde la política, con la discusión de la regulación ética de la IA (Gillespie, 2018).

Pero queda claro que la solución a este problema no solo no es trivial, si no que no estamos ni cerca de resolverlo y por el contrario el problema del contenido basura en internet seguirá en aumento, de ahí la necesidad de reflexionar sobre este tema.

## 2. Taxonomía del contenido basura en el ecosistema digital

La clasificación del contenido inexacto o malintencionado en Internet es fundamental para el desarrollo de estrategias de mitigación precisas.

Existen 5 tipos de información basura que prolifera en internet, como se muestra en la figura 1.

Para algunos de ellos contamos hoy en día con herramientas robustas y eficientes que nos permiten manejar la información, como es el caso del envío de basura por medio de correo electrónico los cuales son clasificados como spam,

**Tabla 2.** Estrategias de mitigación del contenido basura por enfoque

Estrategia	Enfoque Principal	Descripción de la Intervención	Tecnología / Marco Legal
<b>Filtros Automatizados</b>	Tecnológico	Uso de técnicas avanzadas de PLN (Procesamiento del Lenguaje Natural) y <i>Deep Learning</i> para identificar patrones lingüísticos, sintácticos y metadatos de contenido malicioso y <i>spam</i> (Güler & Gündüz, 2023).	Naive Bayes, Support Vector Machines (SVM), BERT, Redes Neuronales.
<b>Verificación de Hechos (Fact-Checking)</b>	Social / Colaborativo	Validación sistemática de la veracidad de contenidos virales por parte de organizaciones independientes. Se integran los resultados para reducir la difusión algorítmica de la falsedad (Johnson, 2024).	PolitiFact, Snopes, Programas de verificación colaborativa en plataformas.
<b>Detección de Desinformación por IA</b>	Tecnológico	Entrenamiento de algoritmos para evaluar la credibilidad de contenidos a partir de coherencia semántica, análisis de fuentes citadas y patrones de difusión (Zhou & Zafarani, 2020).	<i>Machine Learning, Deep Learning, Análisis de Redes.</i>
<b>Marcos Regulatorios</b>	Normativo	Establecimiento de obligaciones claras para las plataformas digitales en materia de transparencia algorítmica y moderación de contenidos. Busca equilibrar la libertad de expresión con la responsabilidad (Klonick, 2017).	<i>Digital Services Act (DSA) de la UE, debate sobre la Sección 230 de EE. UU.</i>
<b>Alfabetización Mediática</b>	Educativo	Desarrollo de competencias críticas en los usuarios para la evaluación de fuentes, el reconocimiento de sesgos cognitivos y el fomento de la conciencia digital (UNESCO, 2022).	Programas educativos, <i>MediaWise, UNESCO MIL CLICKS.</i>

muchas veces de manera transparente para el usuario y en otras permite que lo incorporemos como no deseado o como spam de una manera muy simple. En la tabla 1, se presenta la taxonomía más relevante del contenido basura, haciendo énfasis en la intencionalidad del emisor, un factor clave para su diferenciación.

Adicionalmente a esta clasificación, la IA generativa ha exacerbado la problemática al facilitar la creación automatizada de grandes volúmenes de contenido sin supervisión humana, intensificando la escala y velocidad de la producción de contenido basura (Naples, 2019). Este fenómeno convierte el entorno informativo en

un espacio saturado, y cada vez mas, donde la atención del usuario se convierte en la métrica de valor primario (monetización vía publicidad y datos), un bucle que el contenido basura alimenta constantemente.

### 3. Consecuencias en los ecosistemas social, informativo y tecnológico

El *garbage content* se ha consolidado como un desafío sistémico con ramificaciones, por ejemplo, en lo relacionado con los efectos sociales y de confianza pública: notamos una disminución en la

Confianza, es decir, la desconfianza en los medios e instituciones ha mermado el tejido de credibilidad pública. Estudios indican que solo el 40% de los usuarios confía en las noticias que consume en línea (Newman et al., 2024). También en este mismo rubro tenemos el llamado nihilismo informativo y polarización, el cual se describe como la incapacidad para distinguir entre verdad y falsedad, lo cual lleva a los usuarios a adoptar posturas basadas en afinidades emocionales o ideológicas, profundizando las divisiones sociales (Lazer et al., 2018).

También podemos identificar efectos informativos y cognitivos, tales como la saturación de información o infodemia, la cual implica la sobreabundancia de datos irrelevantes o falsos y dificulta la discriminación de fuentes veraces. Este fenómeno impactó notablemente la gestión de la salud pública durante la pandemia de COVID-19 (Zarocostas, 2020, Cinelli et al., 2020). Además, se presenta la fatiga informativa. Es decir, el exceso de ruido genera una saturación cognitiva que induce a la desconfianza generalizada y al abandono del esfuerzo del pensamiento crítico.

Por último, tenemos los efectos tecnológicos y algorítmicos, por ejemplos los sistemas de recomendación priorizan el *engagement* (interacción) sobre la veracidad, creando un círculo vicioso: el contenido emocional obtiene mayor visibilidad, independientemente de su calidad (Gillespie, 2018), es decir una amplificación algorítmica. Por otro lado, la exposición constante a información dudosa obliga al sistema educativo a integrar la alfabetización mediática e informacional como competencia crítica (UNESCO, 2022).

#### 4. Estrategias multidimensionales de mitigación

El combate efectivo contra el *garbage content* requiere de varias acciones que se deben realizar de manera conjunta tanto tecnológicas, normativas y educativas. El objetivo es construir un ecosistema informativo más resiliente y fiable, en la tabla 2 se muestran algunas de las estrategias que permiten la mitigación del contenido basura en internet, por enfoque.

La regulación busca responsabilizar a las plataformas por el contenido que albergan. Sin embargo, la clave reside en el equilibrio: una sobre-regulación podría derivar en el control gubernamental de la información. La autorregulación de plataformas como Meta y Google es complementaria, aunque su efectividad depende de la coherencia global y la aplicación uniforme de sus políticas de integridad (European Commission, 2022).

La alfabetización digital es el componente indispensable. Ninguna solución tecnológica o regulatoria será suficiente sin usuarios críticos y empoderados. El rol de los usuarios es crucial para la construcción de un ecosistema digital saludable, mediante acciones simples como: reportar contenido falso y no compartir información dudosa.

#### 5. Conclusiones y recomendaciones

La lucha contra el contenido basura en Internet es un imperativo de la sociedad digital. El enfoque debe ser la capacitación ciudadana y el control algorítmico:

- **Prioridad Educativa:** La dotación de habilidades para el análisis crítico, la verificación de fuentes y el reconocimiento de intereses subyacentes es la herramienta más poderosa. La alfabetización mediática debe ser integrada de manera transversal en el currículo educativo en todos los niveles.
- **Autorregulación Digital:** Fomentar prácticas individuales como las pausas informativas, el consumo de fuentes diversas y el uso de listas de confianza para la curación de noticias.
- **Cooperación Institucional:** Los gobiernos y las plataformas tecnológicas deben establecer marcos normativos claros que promuevan la transparencia algorítmica y apliquen sanciones consistentes contra la manipulación informativa.

El avance hacia un ecosistema digital más seguro requiere la colaboración constante entre la academia, el sector público y la sociedad civil para establecer códigos éticos rigurosos que fomenten la responsabilidad informativa y la calidad del dato.

## Referencias

1. **Allcott, H., Gentzkow, M. (2017).** Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, Vol. 31, No. 2, pp. 211–236. doi: 10.1257/jep.31.2.211.
2. **Alao, O., Adepoju, S. (2022).** Machine Learning-Based Fake News Detection on Social Media: Review. *Journal of Information Technology Research*, Vol. 15, No. 2, pp. 1–17.
3. **Cinelli, M., Quattrociochi, W., Galeazzi, A. (2020).** The COVID-19 Social Media Infodemic. *Scientific Reports*, Vol. 10, No. 1, pp. 16598. doi: 10.1038/s41598-020-73510-5.
4. **Corbu, N., Bãrgăoanu, A., Durach, F. (2021).** Fake News Going Viral: The Mediating Effect of Negative Emotions. *Media Literacy and Academic Research*, Vol. 4, No. 2, pp. 58–87.
5. **European Commission. (2022).** The Digital Services Act (DSA). European Commission.
6. **Gillespie, T. (2018).** *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
7. **Goodfellow, I., Bengio, Y., Courville, A. (2016).** *Deep Learning*. MIT Press.
8. **Google AI. (2022).** *Responsible AI Practices*. Google AI.
9. **Güler, G., Gündüz, S. (2023).** Deep Learning Based Fake News Detection on Social Media. *International Journal of Information Security Science*, Vol. 12, No. 2, pp. 1–21. doi: 10.55859/ijiss.1231423.
10. **Hao, K. (2022).** AI Is Making It Easier to Generate Fake News. *MIT Technology Review*.
11. **Klonick, K. (2017).** The New Governors: The Rise of Private Power in the Digital Age. *Harvard Law Review*, Vol. 130, No. 6, pp. 1598–1683.
12. **Johnson, P.R. (2024).** A Case of Claims and Facts: Automated Fact-Checking the Future of Journalism's Authority. *Digital Journalism*, Vol. 12, No. 10, pp. 1461–1484. doi: 10.1080/21670811.2023.2174564.
13. **Lazer, D.M.J., Baum, M.A., Benkler, Y. (2018).** The Science of Fake News. *Science*, Vol. 359, No. 6380, pp. 1094–1096. doi: 10.1126/science.aao2998.
14. **Napoli, P.M. (2019).** *Social Media and the Public Interest: Media Regulation in the Disinformation Age*. Columbia University Press.
15. **Newman, N., Fletcher, R., Schulz, A. (2024).** *Reuters Institute Digital News Report 2024*. University of Oxford.
16. **Pennycook, G., Rand, D.G. (2019).** Fighting Misinformation on Social Media Using Crowdsourced Judgments of News Source Quality. *Proceedings of the National Academy of Sciences*, Vol. 116, No. 7, pp. 2521–2526. doi: 10.1073/pnas.1806781116.
17. **Pennycook, G., McPhetres, J., Zhang, Y. (2021).** Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, Vol. 32, No. 4, pp. 562–574. doi: 10.1177/0956797620974586.
18. **Shu, K., Sliva, A., Wang, S. (2020).** Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, Vol. 19, No. 1, pp. 22–36. doi: 10.1145/3137597.3137600.
19. **Tufekci, Z. (2015).** Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency. *Colorado Technology Law Journal*, Vol. 13, No. 1, pp. 203–218.
20. **UNESCO. (2022).** *Media and Information Literacy Curriculum for Educators and Learners*. UNESCO Publishing.
21. **Vaidhyanathan, S. (2018).** *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford University Press.
22. **Wardle, C., Derakhshan, H. (2017).** *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Council of Europe Report DGI(2017)09.
23. **Wu, T. (2016).** *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Alfred A. Knopf.
24. **Zarocostas, J. (2020).** How to Fight an Infodemic. *The Lancet*, Vol. 395, No. 10225,

pp. 676. doi: 10.1016/S0140-6736(20)30461-X.

*Article received on 18/11/2025; accepted on 09/02/2025.  
\*Corresponding author is Rafael Guzman Cabrera.*

- 25. Zhou, X., Zafarani, R. (2020).** A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, Vol. 53, No. 5, pp. 1–40. doi: 10.1145/3395046.