

# #EcuadorBajoAtaque: Desorden informativo y la difusión de emociones negativas en X durante la crisis de seguridad en 2024

*#EcuadorBajoAtaque: Information Disorder and the Spread of Negative Emotions on X During the 2024 Security Crisis*

*#EcuadorBajoAtaque: Desordem da informação e a disseminação de emoções negativas no X durante a crise de segurança em 2024*

Juan Federico Pino Uribe  
ORCID: 0000-0003-0799-3001

Paula Valentina Zukin Cantor  
ORCID: 0009-0003-5688-9881

Manuel Elkin Castiblanco Briceño  
ORCID: 0009-0000-7940-9594

Juan David Melo Viasús  
ORCID: 0009-0000-8938-751X

DOI: 10.22235/d.v38.4094  
Recepción: 03/06/2024  
Revisión: 29/08/2024  
Aceptación: 29/10/2024

Flacso Ecuador Ecuador.  
Correo: jfpinofl@flacso.edu.ec

**RESUMEN.** Desde 2019, Ecuador ha experimentado un aumento significativo en los homicidios y la expansión de bandas narcotraficantes, lo que alcanzó su punto álgido durante la crisis de seguridad de 2024, marcada por el ataque al canal de televisión TC y atentados en varias provincias del país. En este contexto, el presente estudio tiene como objetivo analizar el desorden informativo (DI) en la plataforma digital X durante dicha crisis a través del análisis de la difusión del DI mediante hashtags y posteos (corpus de 13.603 trinos) de ciudadanos, medios de comunicación y asambleístas, entre el 9 y el 17 de enero. Se empleó una metodología cuantitativa que combina técnicas de procesamiento de lenguaje natural (NLP) y análisis de contenido. Los resultados evidencian la presencia de DI durante la crisis. Se observa una predominancia de emocionalidad negativa en los trinos con DI, mientras que los trinos sin DI mostraron una tendencia hacia emociones más neutras. El estudio subraya cómo las crisis de seguridad facilitan la creación y difusión del DI, destacando la importancia de desarrollar modelos automatizados que permitan su identificación en tiempo real.

**Palabras clave:** desorden informativo; crisis de seguridad; procesamiento de lenguaje natural; Ecuador.

**ABSTRACT.** Since 2019, Ecuador has faced a significant rise in homicides and the growth of narcotics gangs, culminating in the 2024 security crisis marked by attacks on TC Television and violence in several provinces across the country. This study aims to examine the phenomenon of information disorder (ID) on the digital platform X during this crisis by analyzing the spread of ID through hashtags and posts (a dataset of 13,603 tweets) shared by citizens, media outlets, and legislators between January 9 and January 17. A quantitative methodology was applied, combining natural language processing (NLP) techniques with content analysis. The findings reveal the presence of ID throughout the crisis. Tweets containing ID predominantly displayed negative emotional content, whereas tweets without ID tended to reflect more neutral emotions. This study underscores how security crises can fuel the creation and spread of ID, highlighting the need for automated models that enable real-time identification of such content.

**Keywords:** information disorder; security crisis; natural language processing; Ecuador.

**RESUMO.** Desde 2019, o Equador tem experimentado um aumento significativo nos homicídios e a expansão das gangues de narcotráfico, o que atingiu seu ápice durante a crise de segurança de 2024, marcada pelo ataque ao canal de televisão TC e atentados em diversas províncias do país. Nesse contexto, o presente estudo tem como objetivo analisar a desordem da informação (DI) na plataforma digital X durante essa crise, por meio da análise da disseminação da DI mediante hashtags e postagens (corpus de 13.603 publicações) de cidadãos, meios de comunicação e parlamentares, entre 9 e 17 de janeiro. Foi utilizada uma metodologia quantitativa que combina técnicas de processamento de linguagem natural (NLP) e análise de conteúdo. Os resultados evidenciam a presença do DI durante a crise. Observa-se uma predominância de emoções negativas nas postagens com DI, enquanto aquelas sem DI mostraram uma tendência a emoções mais neutras. O estudo destaca como as crises de segurança facilitam a criação e disseminação de DI, destacando a importância de desenvolver modelos automatizados que permitam a sua identificação em tempo real.

**Palavras-chave:** desordem da informação; crise de segurança; processamento de linguagem natural; Equador.

## Introducción

En los últimos años, Ecuador ha enfrentado un incremento en los homicidios y la expansión de bandas armadas vinculadas al narcotráfico (Carrión & Silva, 2023; Reuters, 2023). En este contexto, el 9 de enero de 2024, un grupo de encapuchados perpetró un asalto armado contra el canal TC de Guayaquil. Ese mismo día se reportaron una serie de explosiones en las provincias de Azuay y Guayas, cuyas autorías no fueron esclarecidas, lo que intensificó la percepción de caos e inseguridad. Mientras que algunos sectores de la ciudadanía interpretaron estos ataques como un montaje del Gobierno de Daniel Noboa, otros los consideraron una demostración del poder de los narcotraficantes y la gravedad de la crisis de seguridad en el país. Estas interpretaciones reflejan la influencia del desorden informativo (DI), manifestada en la desconfianza hacia las instituciones del Estado y la proliferación de información falsa, contradictoria y engañosa durante la crisis. Ante esta situación, la ciudadanía recurrió a plataformas digitales en un intento por identificar a los responsables entre las más de veinte bandas armadas que operan en Ecuador (Primicias.ec, 2024).

En respuesta a estos hechos, el presidente Daniel Noboa declaró un estado de excepción tras la fuga de Adolfo Macías, alias Fito, líder de Los Choneros, la mayor banda criminal del país (Pozo Peralta, 2024). Posteriormente, Noboa reconoció la existencia de un conflicto armado interno (Glatsky & Cabrera, 2023). Este anuncio no solo permitió la intervención del ejército para controlar el orden público, sino que también representó un reconocimiento, tanto a nivel nacional como internacional, del desafío que las bandas criminales suponen para el monopolio de la violencia legítima del Estado (Pontón Cevallos & Rivera Velez, 2024).

A pesar de los análisis de la crisis de seguridad realizados por la academia, las organizaciones internacionales, las instituciones públicas y los medios de comunicación, su impacto en el desorden informativo (DI) ha sido menos estudiado (Tufail et al., 2023). Este artículo examina las repercusiones de dicha crisis en el ecosistema informativo digital de Ecuador, planteando la siguiente pregunta: ¿cómo se manifiesta el DI en la plataforma digital X durante la crisis de seguridad iniciada el 9 de enero de 2024?

Este estudio emplea una metodología cuantitativa basada en datos de la plataforma X y se enfoca en el uso de los hashtags más virales y sus niveles de desorden informativo durante la crisis de seguridad. Este tema resulta especialmen-

te pertinente, ya que plataformas digitales como Facebook, Instagram, X y TikTok se han consolidado como espacios clave de interacción política (Manfredi et al., 2024, p. 71).

La literatura destaca que el DI constituye una amenaza para las democracias, ya que socava la confianza en las instituciones. Wardle y Derakhshan (2017) señalan que el DI afecta la percepción pública, fomenta la polarización y desestabiliza los sistemas democráticos (Ramos, 2013). Además de reflejar altos niveles de conflictividad política, el DI facilita la manipulación informativa mediante la construcción de narrativas falsas, exageradas o tendenciosas que engañan a la ciudadanía (Santana Chaves & De Camargo Pentead, 2023). Este fenómeno abarca prácticas como la manipulación mediática, que van desde la difusión de contenidos para promover una visión particular hasta la propagación de información falsa o engañosa con fines de entretenimiento, la promoción de agendas políticas o el ataque a la reputación de adversarios (Marwick & Lewis, 2017). De este modo, el DI, amplificado por las plataformas digitales (Tucker et al., 2018), distorsiona la realidad y contribuye a la polarización política, lo que puede desestabilizar las democracias (Wardle y Derakhshan, 2017).

Ecuador es un caso de estudio relevante porque permite analizar la intersección entre la crisis de seguridad y el DI, aspectos que han adquirido una creciente relevancia a nivel global. La pertinencia de este estudio trasciende las fronteras ecuatorianas, especialmente en un momento en que el aumento de la violencia criminal y las crisis de seguridad se extienden por Latinoamérica, Estados Unidos y Europa (Badrán, 2023; Barrios, 2019). Los resultados pueden contribuir a comprender otros entornos afectados por problemas de seguridad similares. De este modo, el análisis del DI en Ecuador ofrece una contribución significativa a la comprensión de un fenómeno con implicaciones globales.

El artículo se organiza de la siguiente manera: en primer lugar, se realiza la revisión de literatura, seguida de la exposición del marco teórico; a continuación se describe la metodología empleada; y, finalmente, se presentan los hallazgos sobre los patrones de DI observados en X durante la crisis de seguridad en Ecuador.

## Revisión de literatura

El DI ha sido ampliamente abordado en la literatura, especialmente en relación con las elecciones (Allcott & Gentzkow, 2017; Cinelli et al., 2020; Corredoira, 2020;

Gutiérrez-Coba & Rodríguez-Pérez, 2023; Lombana et al., 2022). Sin embargo, existe una carencia de estudios que analicen la difusión de DI tras incidentes de seguridad pública, como la violencia terrorista, el crimen organizado o los conflictos armados, particularmente en contextos fuera de Europa occidental y Estados Unidos (Chen et al., 2022).

Se ha demostrado que los grupos armados también promueven el DI. En el tiroteo de Las Vegas en 2017, un atacante asesinó a 59 personas e hirió a más de 500 en un hotel (Blankstein et al., 2017). El Estado Islámico (ISIS) reivindicó la autoría y afirmó que el atacante era un soldado del califato (Savransky, 2017). No obstante, el FBI refutó esta afirmación y dijo que no existía vínculo entre el atacante y el ISIS (Savransky, 2017). Periodistas señalaron que ISIS era conocido por atribuirse falsamente actos violentos para ganar publicidad (Schwartz, 2017) al difundir propaganda para explotar la inquietud social y reforzar su percepción como una gran amenaza.

Este fenómeno no es aislado, varios autores han identificado que diversos grupos armados utilizan el DI como táctica en conflictos. Al-Shabaab en Somalia y Boko Haram en Nigeria han empleado plataformas digitales para difundir DI con fines de reclutamiento y justificación de ataques (Ortiz dos Santos, 2021; Stremlau, 2022). De manera similar, el Talibán en Afganistán ha utilizado estas plataformas para intimidar a la población y coordinar acciones violentas (Ortiz dos Santos, 2021). Según Piazza (2021), quien estudió este fenómeno en 150 países, el DI es utilizado por grupos terroristas para profundizar la polarización política y el tribalismo dentro de los países.

Investigadores del Centre for Research and Evidence on Security Threats (CREST) también han analizado la influencia rusa en los ataques en el Reino Unido en 2017, donde identificaron el uso de rumores, teorías conspirativas y mensajes difamatorios para polarizar a la sociedad tras un ataque terrorista (Innes, 2017). Johnson (2018) examinó cómo la propagación de noticias falsas y teorías conspirativas se relaciona con la violencia en eventos como *#pizzagate* y el mitin *Unite the Right* en Estados Unidos. Otros estudios rastrearon la difusión de DI en X tras eventos como los ataques en Manchester y el tiroteo de Las Vegas, y observaron un incremento significativo de DI un día después de los incidentes terroristas (Kostakos et al., 2018). Si no se aborda, esta información puede generar inestabilidad y quebrantar comunidades, como ocurrió en Cachemira tras el ataque violento de Pulwama

(Siyech, 2019). Por ello, las autoridades han advertido a la ciudadanía sobre los riesgos de difundir noticias falsas tras una crisis de seguridad (Earley, 2021).

El presente artículo se alinea con los estudios de Chen et al. (2022), quienes examinaron el DI en países del sudeste asiático tras incidentes terroristas que provocaron crisis de seguridad y desarrollaron un algoritmo para su detección. En este marco, el artículo realiza dos contribuciones a la literatura: en primer lugar, mientras que la mayoría de los estudios se han centrado en crisis de seguridad derivadas de actos terroristas, las crisis originadas por el crimen organizado transnacional y su relación con el DI han sido menos exploradas. En segundo lugar, este trabajo avanza en el desarrollo de un modelo automatizado para la detección de DI en crisis de seguridad en América Latina, específicamente en el caso de Ecuador.

### De la desinformación al desorden informativo

La definición del DI ha estado vinculada al concepto de desinformación, que tiene sus orígenes en Rusia en el siglo XX (Razo, 2021). Una de las primeras definiciones del término aparece en el diccionario ruso de Ojegov, donde se describe la desinformación como “la acción de inducir a error mediante el uso de información falsa” (Heller, 1980, p. 167). Desde entonces, este concepto ha ganado relevancia en los estudios de comunicación. Con el auge de las plataformas digitales, estas se han consolidado como fuentes de información para una audiencia global, proporcionando un canal idóneo para la proliferación del DI, incluida la desinformación (Tandoc et al., 2018). Esto se debe a que dichas plataformas permiten a cualquier usuario llegar a un público masivo, sin que la información difundida esté sujeta a procesos de filtrado o verificación.

En las últimas dos décadas del siglo XXI la desinformación ha adquirido mayor relevancia, impulsada por eventos globales como las campañas electorales en Estados Unidos, el Brexit y la pandemia (Melo & Gómez, 2023). Estos escenarios han evidenciado el impacto social que la desinformación puede generar (Taddeo et al., 2022).

El DI también ha estado asociado con el término *fake news* (noticias falsas), que ganó popularidad en el siglo XXI en el debate público. Allcott y Gentzkow (2017) definen las *fake news* como noticias que inducen al engaño y que son intencionalmente y verificablemente falsas. Sin embargo, este término ha sido problemático debido a su

politización (Son & Rashid, 2022). La Comisión Europea ha desaconsejado su uso para referirse a informaciones falsas o erróneas que circulan en internet, ya que no describe adecuadamente la desinformación, que incluye tanto contenido falso como información fabricada o difundida con mala intención (European Commission, 2018). Por ello, se ha propuesto el uso del concepto de DI para capturar mejor los matices de este problema (Taddeo et al., 2022). Según Saló y Pérez-Soler (2022), el DI puede afectar los derechos humanos, como lo demuestran las asonadas contra migrantes en el Reino Unido y España en agosto de 2024.

En este contexto, Wardle y Derakhshan (2017) clasifican el DI en tres tipos: desinformación, misinformación y malinformación. Según estos autores, la desinformación se da cuando se comparte información falsa con la intención de causar daño; la misinformación, cuando se comparte información falsa sin dicha intención; y la malinformación, cuando se comparte información verdadera con el fin de causar daño, a menudo se revelan datos destinados a permanecer privados. Así, la tipología del DI propuesta por Wardle y Derakhshan se basa en si la información es falsa/engañosa y en la intención de causar daño.

Los autores también delinear las fases en las que el DI se desarrolla: creación, donde surge el mensaje; producción, cuando el mensaje se transforma en un producto mediático; y distribución, cuando se vuelve viral. Para facilitar la comprensión del DI, identifican tres elementos clave en su análisis: el agente, que se refiere a quiénes son los responsables de producir y difundir el DI, y cuál es su motivación; el mensaje, que abarca el tipo, formato y características de la información; y el intérprete, que analiza cómo es recibido el mensaje y qué acciones desencadena en el público (Wardle & Derakhshan, 2017).

Este trabajo se enfoca en las categorías de desinformación, malinformación y misinformación (Tabla 1).

Este artículo propone que las crisis de seguridad fomentan la producción de mensajes con DI, ya que facilitan la proliferación de desinformación, misinformación y malinformación. En tiempos de crisis, como desastres naturales, conflictos armados, atentados terroristas y pandemias (Chavero & Intriago, 2021), la urgencia por obtener información lleva a la población a buscar y compartir datos rápidamente, a menudo sin verificar su veracidad. Este fenómeno se agrava por las características de las plataformas digitales, donde la información se

Tabla 1  
Tipología de desórdenes informativos

Categoría	Descripción
Desinformación	La información falsa se difunde intencionadamente para causar daño, a menudo a través de contenidos audiovisuales fabricados o manipulados.
Malinformación	Se comparte información genuina para causar daño, se trasladan datos privados a la esfera pública. Esto incluye la publicación deliberada de información privada por interés personal o corporativo, como la pornografía vengativa o el cambio intencionado de contexto.
Misinformación	Existe información falsa que no tiene la intención de causar daño, como errores involuntarios en pies de foto, fechas, estadísticas, traducciones o cuando la sátira es tomada en serio.

Fuente. Wardle y Derakhshan (2017)

difunde de forma rápida y masiva (Wardle & Derakhshan, 2017). Por lo cual se argumenta que las crisis de seguridad, como la ocurrida en Ecuador en 2024, facilitan la creación y difusión de DI por parte de los usuarios en la plataforma X. El DI es generado y viralizado en este entorno de incertidumbre, aprovechado por los usuarios para compartir mensajes falsos, erróneos o manipulados con mayor rapidez y alcance, ya sea de forma involuntaria o deliberada. Aunque la crisis no tiene agencia, actúa como un catalizador que crea las condiciones para la propagación de este tipo de mensajes.

Por consiguiente, las crisis de seguridad generan un entorno de incertidumbre y miedo, lo que hace que las personas sean más susceptibles a aceptar y difundir información que confirme sus temores y prejuicios (Rodríguez-Ferrándiz, 2023). En situaciones de caos que generan momentos de tensión y estrés, la capacidad de los individuos para discriminar entre fuentes confiables y no confiables se reduce, lo que incrementa la vulnerabilidad a ser engañada por información incorrecta (Lazer et al., 2018; McIntyre, 2018).

Además, las crisis de seguridad son explotadas por actores con agendas que buscan desestabilizar aún más la situación. Estos actores pueden incluir grupos criminales, extremistas políticos o gobiernos extranjeros que desean agudizar las contradicciones políticas. Ellos pueden utilizar tácticas de DI deliberadas para generar confusión, erosionar la confianza en las instituciones y manipular la

opinión pública (Allcott & Gentzkow, 2017; Harsin, 2015). Esta manipulación puede tener consecuencias como el incremento de la violencia, la polarización y la obstaculización de respuestas de emergencia por parte de las autoridades (Guo & Vargo, 2020).

Por lo tanto, este artículo propone que las crisis de seguridad permiten que los individuos produzcan mensajes con DI debido a la combinación de alta demanda de información, la rapidez de difusión en plataformas digitales y el miedo generalizado en la población. Este DI no solo agrava la crisis inmediata, sino que también puede tener efectos a largo plazo en la cohesión social y en la estabilidad institucional al socavar la confianza en las fuentes oficiales y perpetuar un ciclo en el cual el DI y la violencia se refuerzan mutuamente (Rodríguez-Ferrándiz, 2023).

## Metodología

El artículo emplea una metodología cuantitativa que aplica técnicas de procesamiento de lenguaje natural (NLP) para identificar patrones de DI en X durante la crisis de seguridad. Complementariamente, se realiza un análisis cualitativo para ejemplificar el DI presente en trinos, ilustrando cómo actores como la población, los medios de comunicación y los assembleístas participan en este fenómeno. El análisis cuantitativo se basa en la propuesta metodológica de Wan et al. (2022), que ofrece una perspectiva analítica aplicada a estudios de opinión pública mediante el uso de técnicas de NLP.

La metodología consta de cinco etapas. La primera consistió en la recolección de datos en la plataforma digital X, centrada en mensajes relacionados con la crisis de seguridad del 9 de enero de 2024. Esta crisis desencadenó la creación y uso de hashtags para viralizar mensajes sobre la situación (Cañizares et al., 2024). Se utilizaron los hashtags más virales relacionados con la crisis: #EcuadorBajoAtaque, #EcuadorExigeSeguridad, #EcuadorBajoAtaqueTerrorista y #EcuadorEnGuerra. Una vez recolectados los datos, se procedió al preprocesamiento del texto. Este proceso incluyó la estandarización del contenido, la eliminación de caracteres especiales y la lematización<sup>1</sup> del texto, cuando fue necesario, con el objetivo de asegurar que los datos estuvieran en un formato adecuado para su posterior análisis. Durante esta etapa, se calculó la intensidad de las publicaciones realizadas por los usuarios en el período seleccionado, lo cual permitió comprender la frecuencia y distribución temporal de los mensajes. Este análisis fue fundamental para identificar

patrones de comportamiento y picos de actividad relacionados con eventos específicos de la crisis de seguridad. Además, se realizó un análisis cualitativo del contenido de los trinos segmentando los mensajes según el tipo de emisor en población, medios de comunicación y assembleístas para identificar patrones de desinformación (DI) en los trinos.

Para guiar este proceso, el flujo de análisis de datos (*pipeline*) propuesto por Wan et al. (2022) fue adoptado como punto de partida. Sin embargo, dada la particularidad del objetivo, que es identificar y analizar el DI en Ecuador durante la crisis de seguridad de 2024, dicho flujo fue adaptado mediante un ajuste fino (*fine-tuning*) al modelo BERT (Chandra & Saini, 2021; L. Chen et al., 2020; X. K. Chen et al., 2022). Esto permitió clasificar los trinos como DI o no DI. Asimismo, se aplicó un análisis de Keynes para identificar cambios en el uso de palabras en publicaciones con y sin DI. Finalmente, se llevó a cabo un análisis de sentimiento sobre los corpus de DI, lo que permitió evaluar la polaridad de los mensajes e identificar si expresaban emociones positivas, negativas o neutras. Este análisis fue crucial para entender cómo los mensajes de DI manipulan el sentimiento que intentan transmitir a las audiencias virtuales (Wan et al., 2022).

El uso de la plataforma digital X como espacio de estudio resulta adecuado, ya que, a pesar de contar con un menor número de usuarios en comparación con otras plataformas como Facebook o TikTok (Abad, 2017), ofrece un punto de convergencia, confrontación y politización para diversas ideas y discursos (Argüello Pazmiño & Hurtado Arroba, 2024). Esto facilita el análisis de cómo los medios de comunicación, los líderes políticos y la ciudadanía interactúan. Este entorno ofrece datos sobre la dinámica de la formación de opiniones y la difusión de información (Cifuentes, 2018). Además, X se caracteriza por ser un espacio donde los líderes políticos y los medios de comunicación están activos (Abadía et al., 2023; Gómez et al., 2022). Los líderes políticos utilizan este medio para interactuar con sus seguidores y antagonistas, lo que permite estudiar las interacciones y respuestas de las audiencias (Gómez et al., 2022; Manfredi & González, 2019). Este comportamiento contrasta con el uso predominantemente orientado al entretenimiento en

---

<sup>1</sup>: La lematización es un proceso en el análisis de lenguaje natural (NLP) que permite reducir las formas variantes de una palabra a su forma base o lema. Este proceso mejora la precisión de tareas de procesamiento de lenguaje natural, como la clasificación de texto, el análisis de sentimientos y la extracción de información (Khyani et al., 2021).

plataformas como Facebook y TikTok, donde el contenido es más lúdico y menos enfocado en la discusión política (McClain et al., 2024).

### Recolección de datos de plataformas digitales

La recolección de datos se llevó a cabo en X empleando técnicas de *web scraping* mediante Python. Se utilizó la búsqueda avanzada de X para lo cual se desarrolló un *script* en Python utilizando Selenium (<https://www.selenium.dev/>), lo que facilitó la extracción automatizada de los datos. Inicialmente, se extrajo información relacionada con las siguientes palabras clave: Ecuador, terrorismo, narcoterrorismo, narco, violencia, seguridad, ataque, Ecuador bajo ataque. También se utilizó la expresión “Ecuador Bajo Ataque OR Ecuador AND Terrorismo OR Narcoterrorismo OR Narco OR Violencia OR seguridad OR ataque. En esta primera iteración se recopilamos 1.100 trinos, que permitieron ampliar la búsqueda de información al identificar hashtags relevantes, y a partir de estos recopilar más información.

Tabla 2  
Hashtags en el marco de la crisis de seguridad

Hashtags	Cantidad
URGENTE	519
EcuadorBajoAtaque	460
Ecuador	229
Urgente	206
EcuadorExigeSeguridad	139

La segunda iteración de recopilación de información uso la expresión #EcuadorBajoAtaque OR #EcuadorExigeSeguridad, esta permitió recopilar 2.952 trinos publicados entre el 9 y el 17 de enero de 2024, estos datos permitieron identificar otros hashtags utilizados durante la crisis de seguridad en Ecuador.

Tabla 3  
Priorización de hashtags en el marco de la crisis de seguridad

Hashtags	Cantidad
#EcuadorEnGuerra	102
#EcuadorBajoAtaqueTerrorista	101
#EcuadorExigeSeguridad	88

Estos nuevos hashtags complementaron la expresión de búsqueda #EcuadorBajoAtaque OR #EcuadorExigeSeguridad OR #EcuadorBajoAtaqueTerrorista OR #EcuadorEnGuerra, recopilando 11.498 trinos en total. Adicional a estos trinos se recolectó información de 97 cuentas de asambleístas<sup>2</sup> durante el período comprendido entre el 9 y el 17 de enero de 2024, se obtuvo un total de 2.105 trinos.

### Preprocesamiento de texto

El preprocesamiento de datos para este artículo se dividió en dos fases. La primera consistió en el etiquetado manual de los conjuntos de datos seleccionados para entrenar el modelo que identifica la presencia de DI en los trinos. Se trabajó con tres conjuntos de datos: uno correspondiente a la población, otro a los asambleístas y, el último, a los medios de comunicación. Este proceso incluyó la identificación manual de las cuentas de los medios y la selección de los trinos publicados por los asambleístas durante la crisis de seguridad.

Tabla 4  
Relación trinos descargados y etiquetados

Actor	Trinos descargados	Trinos etiquetados*	N.º de usuarios
Asambleístas	2.105	423	97
Medios de comunicación	496	496	22
Población civil	11.002	5.629	5.395
Total	13.603	6.548	5.514

\* Para el etiquetado solo se utilizaron trinos que tuvieran como tema la crisis de seguridad del Ecuador

En esta etapa se utilizó la herramienta Doccano (Nakayama et al., 2018) para etiquetar un total de 6.548 trinos: 423 de asambleístas, 496 de medios de comunicación y 5.629 de la población civil. Cada trino fue leído y clasificado como DI siguiendo las categorías de Wardle y Derakhshan (2017) y los criterios específicos detallados en el Apéndice A1. Aquellos trinos que no presentaban indicios de DI fueron etiquetados como “No DI”. Estos criterios permitieron una clasificación más precisa de los trinos al identificar la intencionalidad detrás de la información compartida, así como

2:: No fue posible recopilar información de los 137 integrantes de la Asamblea Nacional del Ecuador debido a que no todos los asambleístas tienen una cuenta en X, o no realizaron publicaciones durante la crisis de seguridad.

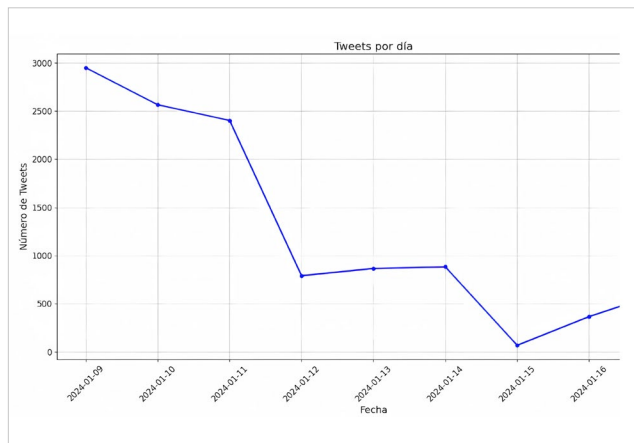
posibles errores involuntarios o manipulaciones deliberadas de contenido. Posteriormente, se filtraron y eliminaron los trinos escritos en idiomas distintos al español, aquellos que no trataban temas relacionados con la crisis de seguridad en Ecuador y los que solo contenían emojis, ya que fueron considerados como ruido en el análisis.

En el segundo momento, se implementaron técnicas para la limpieza y preparación de los trinos con el objetivo de optimizar los análisis (análisis de sentimientos, identificación de desórdenes informativos, análisis de Keynes). Esto incluyó la eliminación de elementos no deseados como menciones, hashtags y caracteres especiales que podrían distorsionar los análisis.

## Resultados

Esta sección identifica la actividad en X durante la crisis de seguridad en Ecuador, destacando los trinos de la población, asambleístas y medios de comunicación a lo largo del tiempo. Este análisis muestra las dinámicas de participación y la respuesta en X durante y tras la crisis.

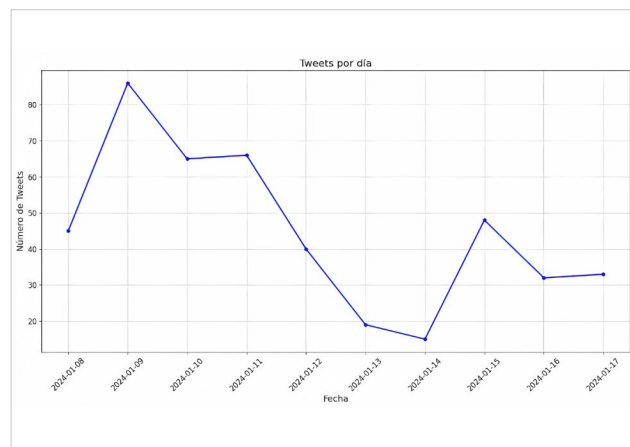
Figura 1  
Trinos de población



La Figura 1 muestra la cantidad de trinos durante la crisis de seguridad en Ecuador. La actividad alcanzó un pico de 3.000 trinos el 9 de enero de 2024, que coincide con el asalto al canal TC y la violencia subsiguiente. Luego descendió a 1.500 el 10 de enero y 500 el 12 de enero. Hubo una estabilización en 400 trinos el 13 de enero, seguida por una caída a 200 el 14 de enero y un ligero aumento a 300 el 17 de enero. Esto sugiere que

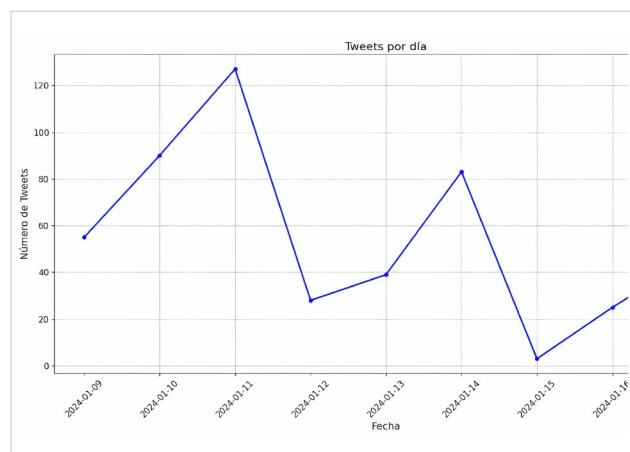
la atención y participación en X fue alta al inicio de la crisis, pero disminuyó gradualmente.

Figura 2  
Trinos de asambleístas



La Figura 2 muestra los trinos de los asambleístas. El 9 de enero de 2024 hubo cerca de 70 trinos, se alcanzó un máximo de 80 el 10 de enero, lo que refleja una respuesta inmediata a la violencia. A partir del 11 de enero, los trinos disminuyeron a 30 el 13 de enero, con una leve recuperación a 40 el 15 de enero. Posteriormente, la actividad fue irregular y baja, con 20 trinos el 17 de enero. Esto sugiere que, aunque inicialmente activos, la participación de los asambleístas disminuyó gradualmente.

Figura 3  
Trinos de medios de comunicación



La Figura 3 muestra los trinos diarios de los medios de comunicación. El 9 de enero de 2024 se registraron 120 trinos, que bajaron a 50 el 10 de enero, subieron a 100 el 11 y cayeron a 40 el 12 de enero. Luego, fueron fluctuantes: 70 trinos el 14 de enero, 20 el 16 y 30 el 17. Este patrón refleja una actividad variable y una cobertura continua en las plataformas digitales.

En las Figuras 1, 2 y 3 se observa un incremento de la conversación a partir del 14 de enero de 2024. Para comprender esta dinámica, se analizó el contexto dentro de este rango de fechas y se examinó si tenía relación con el aumento de trinos; a pesar de haberse identificado previamente un cese acelerado en la interacción en la red social. Algunos hechos destacados son:

- Publicación constante de comunicados oficiales por parte de las Fuerzas Armadas (FFAA) resaltando los resultados, en su mayoría cuantitativos, de los operativos que se realizaron en diferentes zonas y cárceles del país para retomar el orden público. Estos resultados consistían en: liberación de secuestrados, aprehensión de presuntos integrantes y cabecillas de bandas delincuenciales, incautación de armamentos, sometimiento de presos (Página 12, 2024).
- Informe por parte de la agencia penitenciaria sobre la fuga de más de 40 presos en una de las cárceles (El Economista, 2024).
- Diferentes medios de comunicación citaron estudios de varias organizaciones que situaban a Ecuador como uno de los países con índices de violencia más altos, comparándolo con Siria, Irak y Afganistán, y en Latinoamérica dejándolo entre los últimos lugares en el ranking de seguridad (Noboa Gonzalez, 2024).
- Declaración de la Unicef donde afirmaba el incremento desmedido de la tasa de homicidios contra niños, niñas y adolescentes (France 24, 2024).
- Asesinato de Cesar Suárez, fiscal que integraba el equipo de investigación del asalto al canal TC (BBC News Mundo, 2024).
- Pronunciamiento por parte de familiares y abogados de presos expresando la preocupación por la integridad y protección de los derechos humanos de las personas privadas de libertad en medio de los operativos (Navarrete, 2024).

- Envío del proyecto de ley con carácter de urgencia por parte del presidente Noboa para poder enfrentar el conflicto armado interno; dentro del cual se destacó el incremento del IVA para poder financiar la “guerra”. A partir de ese documento, los legisladores pusieron en discusión su apoyo u oposición a dicho proyecto (Salazar, 2024).

### Mapeo de frecuencias del DI

A continuación, se realizó un análisis en nubes de palabras basado en un modelado de tópicos aplicado a los trinos emitidos por medios de comunicación, población civil y asambleístas en la plataforma X, entre el 9 y el 17 de enero de 2024. El modelado de tópicos permitió identificar patrones temáticos y diferencias lingüísticas entre los trinos que contienen DI y aquellos que no, y mostró cómo estos últimos ofrecían más información y mayor claridad en los discursos narrativos de los actores. Los resultados revelan cómo cada grupo contribuye a la narrativa de la crisis, destacando términos que amplifican la percepción de inseguridad.

Al analizar las nubes de palabras de los trinos publicados por los medios de comunicación (Figura 4), se observa una clara distinción entre aquellos que contienen y los que no contienen DI. En los trinos sin DI, predominan términos como “urgente”, “Ecuador”, “presidente”, “Daniel Noboa” y “seguridad”, lo que sugiere una narrativa centrada en la cobertura institucional y política, lo que resalta la importancia de la seguridad y la urgencia en las comunicaciones oficiales. En contraste, en los trinos con DI se identifican términos como “Ecuador”, “terrorista”, “militares”, “estado de guerra”, “narco” y “atención”, asociados a un lenguaje que enfatiza el conflicto, la violencia y la presencia de actores armados. Esta narrativa contribuye a amplificar la percepción de caos e inseguridad en el país. Además, el uso recurrente de términos relacionados con el terrorismo y actividades delictivas refuerza la sensación de crisis y desorden.

La Figura 5 presenta el análisis de los trinos de la población con DI y revela una narrativa emocional y agresiva, centrada en términos como “terrorista”, “delincuente” y “guerras”. Además, en el procesamiento de los datos se evidenció la “petición” generalizada de medidas extremas contra criminales y presos, anulando



Figura 4  
Nube de palabras trinos de medios de comunicación

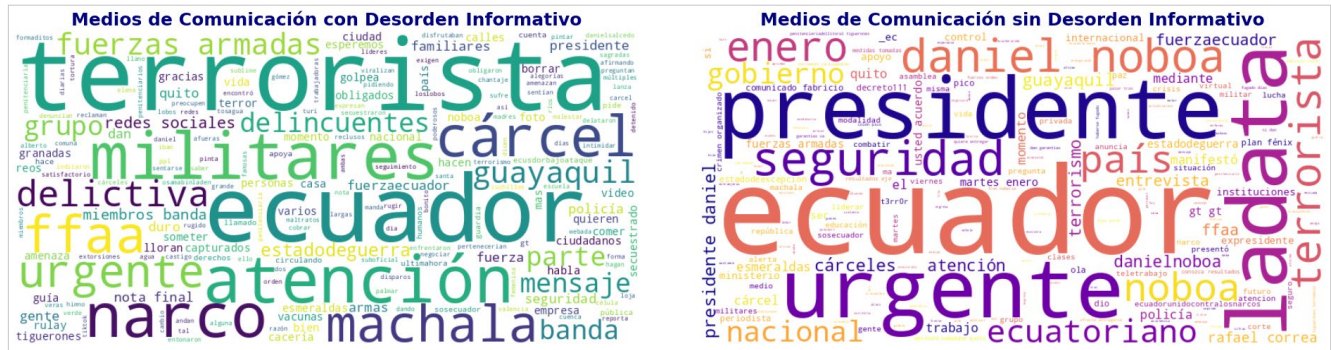
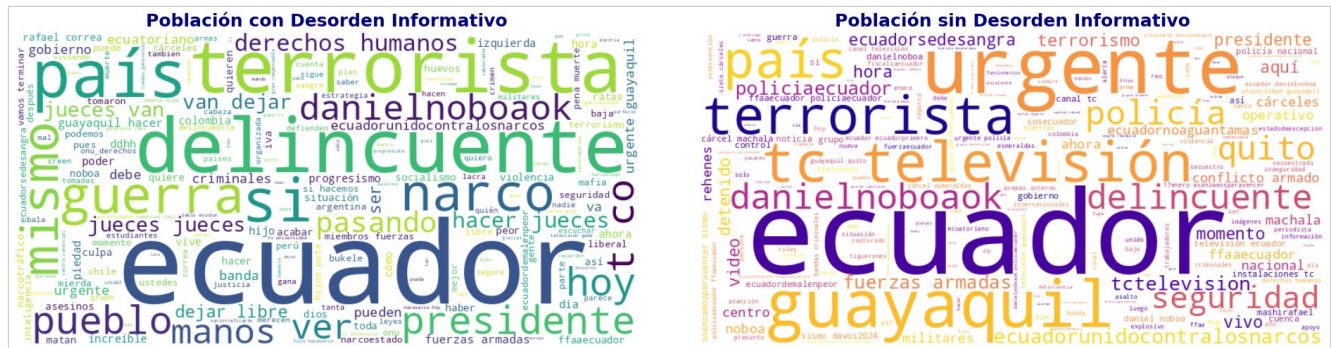


Figura 5  
Nube de palabras trinos de población



la perspectiva de DDHH. Este lenguaje permite intuir una sensación de caos y desesperación ante la crisis. En contraste, los trinos de la población sin DI se enfocan en la réplica de la información mediática y las respuestas institucionales, con términos como “urgente”, “televisión” y “policía”, lo que muestra una narrativa orientada a informar la acción organizada por parte de las autoridades.

En este caso, el desorden informativo refleja caos, mientras que los trinos sin DI adoptan un enfoque más estructurado, menos alarmista y buscan contribuir con la viralización de la información.

En las nubes que presenta la Figura 6 se puede evidenciar que la nube que contiene DI (izquierda) usa términos como “gobierno”, “país”, “seguridad”, “crimen organizado”, “delincuentes” y se centra en ser crítico ante el gobierno y la situación de inseguridad del país. Palabras como “delincuentes”, “mafias” y “crimen orga-

nizado” resaltan un discurso con foco en la criminalidad. También se mencionan con frecuencia figuras políticas, como Lasso y Moreno, lo que sugiere un enfoque en la responsabilidad de los líderes políticos respecto a la crisis. Por otro lado, la nube sin DI tiene un enfoque más positivo con palabras como “paz”, “vida”, “unidad” y “apoyo”; también se destacan términos relacionados con medidas para resolver los problemas de inseguridad, como “acciones”, “momento” y “urgente”; lo que deja entrever un tono más propositivo para resolver la crisis.

El análisis de contenido realizado de forma manual revela una comunicación aparentemente “imparcial”, centrada en informar más que en generar opinión. No obstante, se identificaron trinos que destacan por su potencial tendencia a crear prejuicios contra ciertos individuos o colectividades, como el publicado por *La Data Ec* en relación con un nuevo requisito para ingresar a la Policía Nacional.

Figura 6  
Nube de palabras trinos de asambleístas



Figura 7  
Nube de palabras texto etiquetado como no desorden informativo



Este tipo de trinos puede convertirse en narrativas de discriminación basadas en estereotipos, por lo que es pertinente cuestionar si el simple acto de informar sin cuestionar la veracidad de la información publicada podría contribuir indirectamente al DI y procesos de estigmatización contra grupos poblacionales como las personas con tatuajes (Figura 8)

Para comprender el desorden informativo en función de estas expresiones, se utilizó un modelo de clasificación de sentimientos.<sup>3</sup> El modelo de beto-sentiment-analysis<sup>4</sup> (Pérez et al., 2024) identifica si un sentimiento es positivo, negativo o neutro analizando el contexto y las palabras en una oración de manera bidireccional. Este enfoque permite captar el significado completo de los términos al tener en cuenta las palabras tanto antes como después de cada una. No obstante, la clasificación de sentimientos neutros resulta más compleja, ya que estas expresiones suelen reflejar tonos descriptivos sin carga emocional evidente, como “normal” o “correcto”. Aunque eficaz en la identificación de patrones lingüísticos, BETO enfrenta limitaciones significativas para capturar matices como el sarcasmo, la ironía o el desacuerdo,

especialmente en contextos donde no necesariamente se expresan emociones negativas.

3:: la detección de polaridad de manera explícita y descripción de cómo los datos se clasifican en positivos, negativos o neutros. La detección de polaridad se define como la tarea de predecir si un texto, como un trino, tiene un sentimiento general positivo, negativo o neutral. Esta es la forma estándar en la que se aborda la tarea de análisis de sentimientos o polaridad. A través del ajuste fino, el modelo aprende a captar estas diferencias, asignando la categoría correcta según las probabilidades de cada sentimiento (Pérez et al., 2024).

4:: Este modelo es un *fine-tuning* de BETO: Spanish BERT, un modelo basado en la arquitectura Transformers. Esta arquitectura procesa secuencias de datos mediante capas que integran mecanismos de autoatención y redes de retroalimentación. Cada *token* se convierte en una representación numérica con una codificación de posición específica. La autoatención de múltiples cabezas analiza las relaciones entre palabras, mientras que el apilado de capas con conexiones residuales facilita un aprendizaje más estable y efectivo. Su mecanismo de autoatención tiene potencial para el análisis de sentimientos al capturar de manera efectiva las relaciones contextuales en los textos. Esto permite identificar el sentimiento de manera más precisa, ya que el significado emocional de una palabra suele depender del contexto (Cañete et al. 2020; Pérez et al., 2024).

Figura 8



Fuente. La Data Ec (2024)

Según Mao et al. (2024), los enfoques de análisis de sentimientos basados en modelos como BETO pueden verse afectados por la falta de un análisis contextual más profundo, lo que podría generar errores en la clasificación de emociones o en la interpretación de estados no emocionales. En este sentido, en contextos de participación democrática, el desacuerdo no debería ser clasificado automáticamente como una emoción negativa, ya que representa un componente natural del debate argumentativo. Para mitigar estas limitaciones, sería conveniente que futuros estudios implementen enfoques híbridos y técnicas de refinamiento contextual a fin de mejorar la precisión del análisis, especialmente en dominios y contextos lingüísticos diversos. Estas mejoras permitirían una mayor adaptabilidad del modelo a situaciones complejas, evitarían la clasificación errónea de sentimientos y asegurarían un análisis más contextualizado.

El modelo evidenció (Figura 9) la polaridad emocional en los asambleístas, los medios y la población en los trinos tanto con DI como sin él. La gráfica titulada “Clasificación de polaridad de emociones con Desorden Informativo” muestra que en los trinos con DI predominan los sentimientos negativos, seguidos por neutros y en menor medida positivos. Por otro lado, en la gráfica titulada “Clasificación de polaridad de emociones sin Desorden Informativo” los trinos tienden a ser mayormente neutros, con una menor presencia de emociones negativas. Esto permite identificar que los trinos con DI buscan generar emociones negativas en los receptores en el corpus estudiado.

El análisis de la Figura 9 evidencia la prevalencia de emociones con polaridad negativas en los mensajes emitidos por los actores analizados, los cuales han sido clasificados bajo el concepto de DI. Los resultados presentados

en la Figura 10 subrayan que la indignación constituye la emoción predominante y desempeña un papel importante en la amplificación mediática del DI (Crespo-Martínez et al., 2022).

Figura 9  
Clasificación de polaridad de emociones

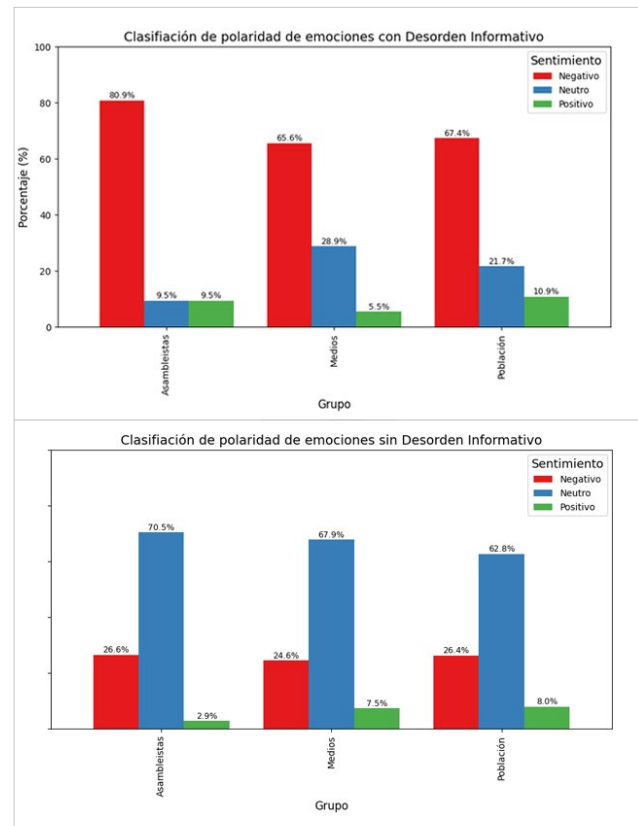
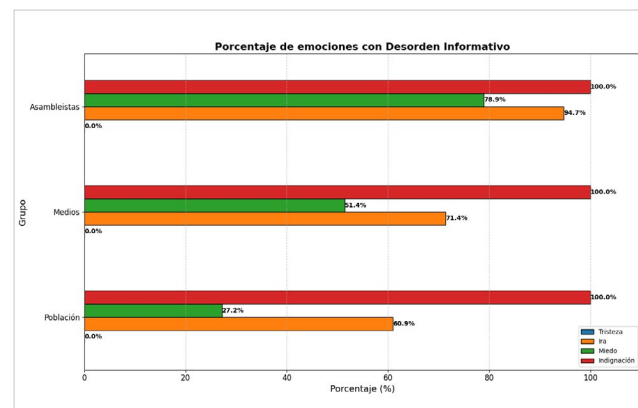


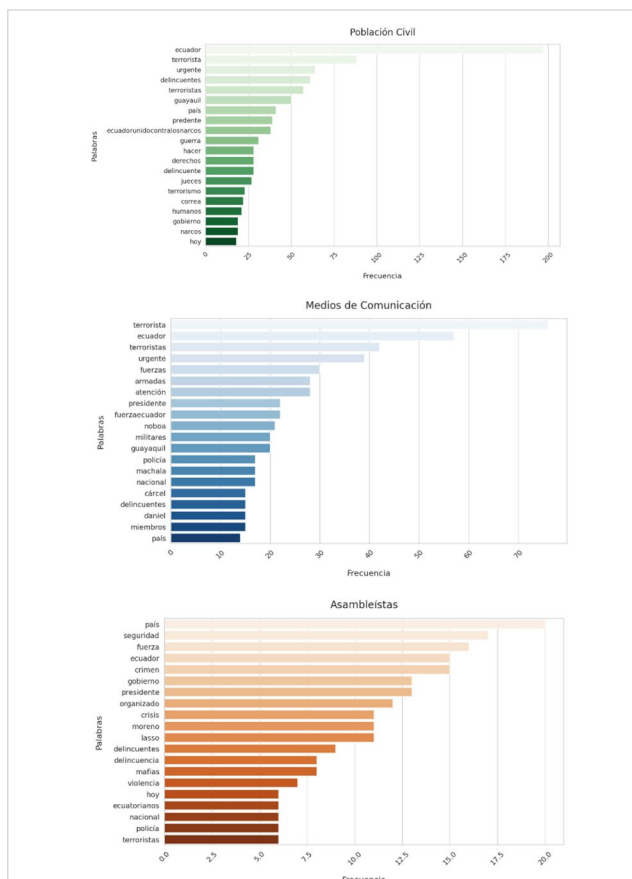
Figura 10  
Clasificación de emociones por actores



Los niveles más elevados de indignación, ira y miedo se manifiestan en los asambleístas, seguidos por los medios de comunicación y, en menor medida, por la población civil. Esto sugiere que el DI surge como una reacción ante una emergencia pública de alto interés como las crisis de seguridad. Los asambleístas, como figuras de autoridad y credibilidad, emiten trinos que pueden basarse en información oficial, extraoficial o en sus propias interpretaciones. Estos trinos son amplificados por los medios de comunicación, que gozan de la confianza y el alcance necesario para llegar a un público amplio. La población, al recibir esta información junto con la de otras fuentes, tanto oficiales como no oficiales, queda expuesta a un flujo constante de datos sin filtrar, lo que puede alimentar narrativas de miedo e indignación.

Los gráficos de barras de la Figura 11 ilustran la frecuencia de palabras en textos que contienen DI. Entre la población, términos como “Ecuador”, “terrorista”, “urgente”

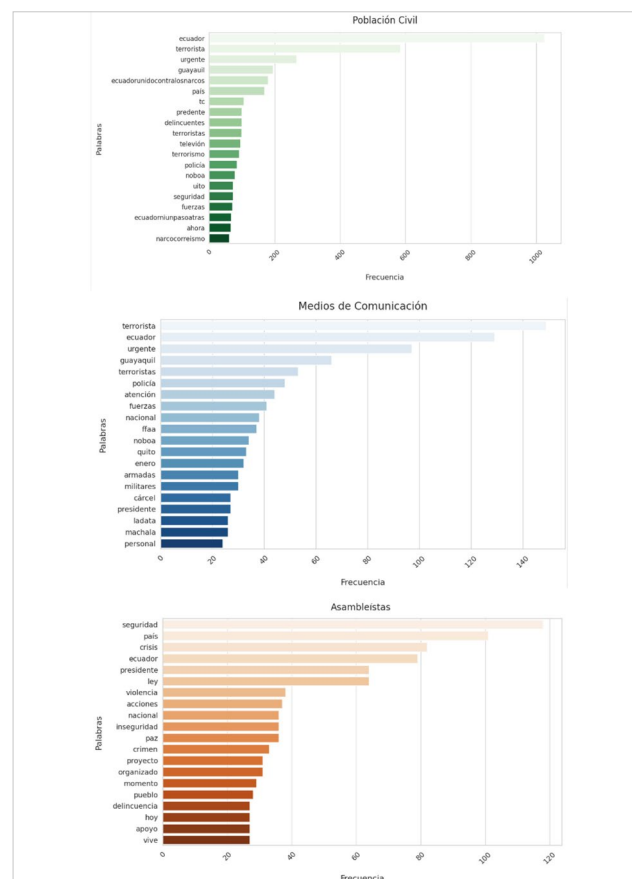
Figura 11  
Frecuencia de palabras texto etiquetado como DI



y “delincuentes” prevalecen, esto refleja una preocupación central por la seguridad y la criminalidad. Los medios de comunicación también subrayan “terrorista”, “Ecuador” y “urgente”, además de destacar a las “FFAA”, lo que indica una narrativa orientada a la respuesta institucional coercitiva. Los asambleístas, por su parte, recurren con frecuencia a “país”, “seguridad” y “fuerza”, se concentran en cuestiones de seguridad y políticas gubernamentales. En conjunto, estos gráficos revelan un enfoque común en la seguridad nacional y la respuesta a las amenazas mediante vías coercitivas.

En general, los textos con discurso incendiario emplean un lenguaje más alarmista y repetitivo al abordar la crisis de seguridad y sus características. En cambio, los textos sin dicho enfoque (Figura 12) tienden a mantener una discusión más equilibrada, centrada en la institucionalidad y la implementación de políticas.

Figura 12  
Frecuencia de palabras texto etiquetado como no DI

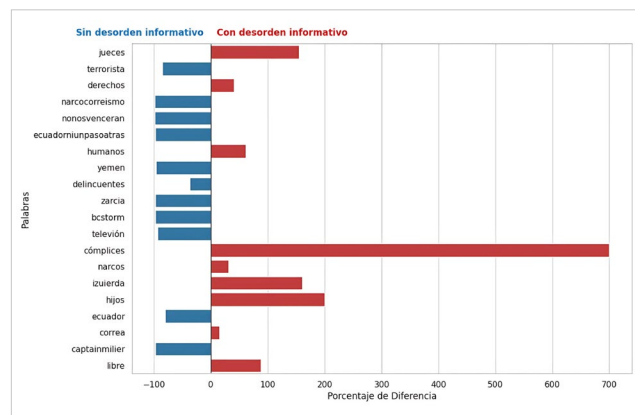


## Análisis de Keynes

El análisis de Keynes es una herramienta que permite identificar y comparar diferencias significativas en la frecuencia de palabras entre distintos conjuntos de textos. En este artículo se aplica para analizar los trinos con y sin DI durante la crisis de seguridad en Ecuador en 2024, lo que facilita observar las variaciones temáticas dentro de cada grupo y comparar cómo estas contribuyen al uso del DI.

La Figura 13 muestra diferencias en la frecuencia de palabras entre los textos con y sin DI en la población civil. En los textos con DI, destacan palabras como “jueces”, “cómplices”, “narcos”, “izquierda”, “hijos”, “Correa libre” y “derechos humanos”. Términos como “narcos”, “izquierda”, “cómplices” y “Correa”, empleados por la población, tienden a resaltar temas controversiales en Ecuador y a responsabilizar a políticos de izquierda, como Rafael Correa. Esto se debe a que en varios trinos se afirmaba que decisiones del expresidente Correa, como la de expulsar a las fuerzas de Estados Unidos al no renovar el contrato de la base militar en Manta, facilitaron la expansión del narcotráfico en Ecuador (Glatsky & Cabrera, 2023; Navarro, 2021).

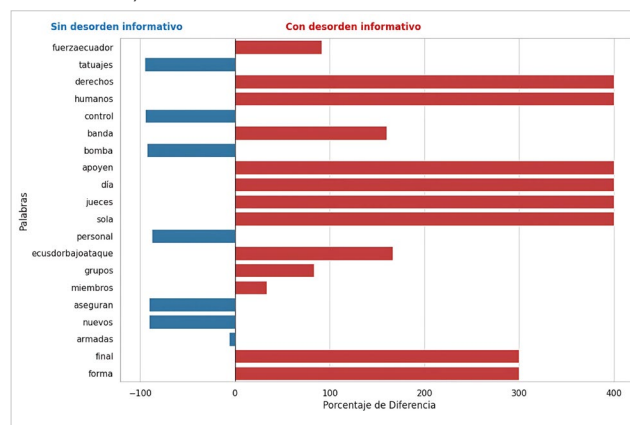
Figura 13  
Análisis de Keynes en la población



El análisis de la Figura 14 para los medios de comunicación revela una clara divergencia en el uso de términos entre los trinos con y sin DI. Las palabras asociadas con el DI, como “tatuajes”, “derechos humanos”, “bomba”, “jueces” y “nuevos” son predominantes, lo que sugiere un énfasis en temas negativos. Estos términos están orientados a generar alarma, utilizar prejuicios y destacar aspectos controvertidos de la crisis. En contraste, los trinos sin DI emplean palabras como “fuerzaEcuador” y “personal”, y se enfocan más en

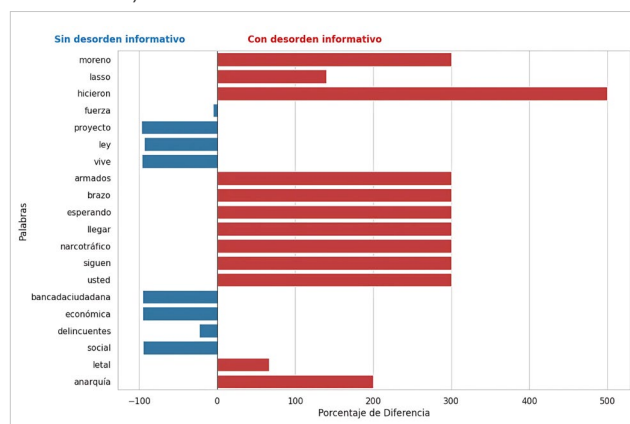
el control y las acciones institucionales, lo que sugiere una narrativa más equilibrada y menos alarmista. Esta dicotomía en el uso del lenguaje refleja cómo los medios pueden influir en la percepción pública de la crisis, potenciando el DI y la sensación de inseguridad.

Figura 14  
Análisis de Keynes en los medios de comunicación



El gráfico de la Figura 15 destaca una significativa diferencia en el uso de términos entre los trinos con y sin DI. Las palabras predominantes en los trinos con DI, como “Moreno”, “hicieron”, “brazos”, “narcotráfico” y “letal”, reflejan un enfoque en figuras políticas, eventos dramáticos y narrativas de conflicto y violencia. Esto sugiere un intento de polarizar la opinión pública y utilizar el contexto de crisis para atacar o defender posiciones políticas. Por otro lado, los trinos sin DI incluyen términos como “proyecto”, “bancadaCiudadana” y “social”, lo que indica un enfoque

Figura 15  
Análisis de Keynes en los assembleístas



más orientado a propuestas y soluciones legislativas. Este contraste evidencia cómo los assembleístas pueden contribuir al DI mediante el uso de un lenguaje alarmista y centrado en la confrontación.

Los gráficos de Keynes para la población civil, los medios de comunicación y los assembleístas durante la crisis de seguridad revelan cómo cada grupo contribuye al DI. La población utiliza términos controversiales y culpabilizadores, los medios de comunicación se enfocan en temas alarmistas y los assembleístas destacan figuras políticas y eventos dramáticos. En contraste, los trinos sin DI en estos grupos presentan un lenguaje menos alarmista. Esta comparación muestra que, mientras la población y los assembleístas tienden a polarizar el discurso, los medios amplifican la percepción de inseguridad.

#### Modelo de clasificación de DI

Para automatizar la identificación de DI en crisis de seguridad, siguiendo la metodología de Chen et al. (2022), se realizó un ajuste fino (*fine-tuning*) en la versión en español de BERT (bert-base-spanish-wwm-uncased). Esta tarea exigió comprender las variaciones del lenguaje en contextos de crisis política, lo que implicó varias iteraciones para mejorar la efectividad del modelo. Las etiquetas de los textos con DI se transformaron en valores numéricos para facilitar su interpretación y reducir el costo computacional para manejar grandes volúmenes de datos.

El conjunto de datos se dividió en un 80 % para entrenamiento y un 20 % para prueba, lo que validó la capacidad del modelo de generalizar en datos no vistos. A partir de técnicas de sobremuestreo, se balanceó el conjunto de datos logrando 4.500 ejemplos por cada clase (DI y no DI). Durante el entrenamiento se ajustaron los hiperparámetros y se añadieron capas de dilución (*dropout*) para prevenir el sobreajuste. Esto mejoró la capacidad del modelo para diferenciar entre narrativas alarmistas y discusiones equilibradas, lo que permitió una comprensión más precisa de cómo se manipula la opinión pública mediante el lenguaje en X.

El modelo fue validado con un conjunto de 1.812 datos etiquetados, con el objetivo de verificar su capacidad para identificar correctamente los textos que contienen DI. Los resultados de esta validación se presentan en la Tabla 5.

Tabla 5  
Reporte de clasificación del modelo de BERT ajustado

	Precisión	Recall	F1-Score	Soporte
DI	0.892057	0.972253	0.930430	901
No DI	0.969880	0.883644	0.924756	911
Accuracy	0.927704	0.927704	0.927704	0.927704
Promedio macro (Macro avg)	0.930968	0.927949	0.927593	1.812
Promedios ponderados (Weighted avg)	0.931183	0.927704	0.927577	1.812

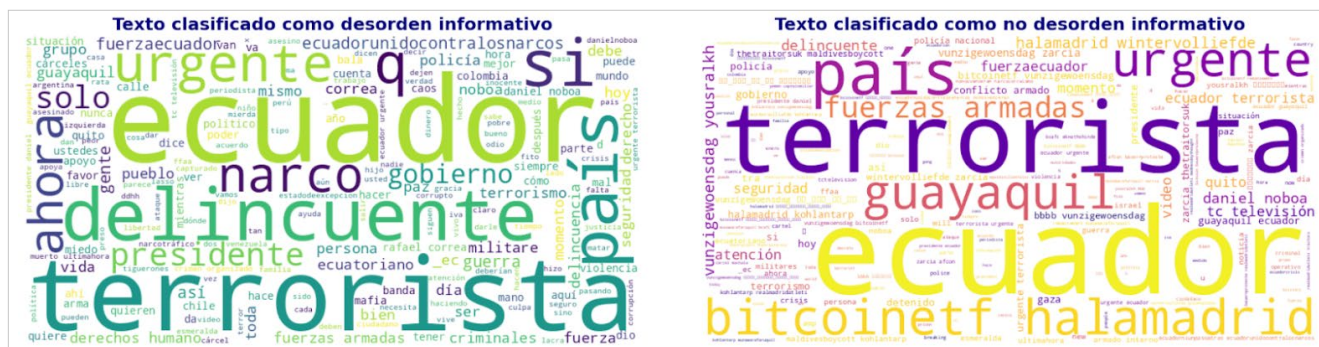
El modelo de clasificación demostró un buen rendimiento en la detección de textos con y sin DI. Para la clase DI, el modelo alcanza una precisión de 0.8921, lo que indica una baja tasa de falsos positivos, y un *recall* alto de 0.9723, lo que sugiere que el modelo identifica correctamente la gran mayoría de los ejemplos con DI. El F1-Score de 0.9304 refleja un excelente equilibrio entre precisión y *recall*, lo que indica un rendimiento altamente efectivo para esta clase.

En cuanto a la clase sin DI, el modelo muestra una precisión muy alta de 0.9699, lo que implica que casi todos los ejemplos etiquetados como No DI son correctos. Su *recall* de 0.8836 sugiere la presencia de algunos falsos negativos, aunque en menor medida. El F1-Score de 0.9248 refleja un excelente equilibrio entre precisión y *recall* también para esta clase.

La exactitud global del modelo es del 92.77 %, lo que indica un rendimiento general sólido. Además, los promedios ponderados muestran una alta precisión (0.9312) y un F1-Score robusto (0.9276), lo que sugiere que el modelo tiene un desempeño equilibrado en ambas clases, interpreta correctamente los textos con y sin DI. Esto representa una mejora significativa respecto a evaluaciones anteriores, destacando la capacidad del modelo para identificar con precisión los textos con y sin DI.

El modelo de clasificación muestra un desempeño adecuado al diferenciar entre textos con potencial de desorden informativo y aquellos que no lo presentan. La nube de palabras correspondiente a los trinos con DI resalta términos que suelen despertar emociones intensas como miedo, inseguridad, ira e indignación. Palabras como “terrorista”, “delincuente” y “narco” dominan el análisis, junto a referencias a “Ecuador”, lo que refuerza la idea de un contenido alarmista y polarizador, característico del DI.

Figura 16  
Nube de palabras de texto clasificado por el modelo de población y medios de comunicación



## Conclusiones

El artículo abordó la identificación del DI en X durante la crisis de seguridad en Ecuador que comenzó el 9 de enero de 2024. Los resultados mostraron una presencia diferenciada de DI en los trinos de población, asambleístas y medios, caracterizada por la predominancia de emociones negativas. La indignación representó el 100 % de las emociones en todos los grupos (asambleístas, medios y población). El miedo fue la segunda emoción más frecuente, con el 94.7 % en los asambleístas, el 71.4 % en los medios y el 60.9 % en la población civil. La ira también fue relevante entre los asambleístas (78.9 %), pero menor en la población (27.2 %). No se detectó tristeza. Estos resultados destacan cómo la crisis amplificó tanto el DI como la carga emocional en los mensajes y acentuó la percepción de inseguridad.

Desde una perspectiva empírica, el estudio realiza contribuciones al identificar patrones recurrentes en los trinos con DI durante la crisis en Ecuador. Términos como “terrorista”, “narcotráfico” y “crisis” dominaron estos mensajes, que refuerzan el vínculo entre el uso de lenguaje alarmista y la proliferación del DI en situaciones de crisis de seguridad. El análisis de las emociones subyacentes evidenció que la indignación, como emoción predominante, está recurrentemente presente en los trinos con DI. Este hallazgo no solo resalta la importancia de las emociones negativas en los trinos con DI, sino que también subraya la necesidad de considerar el componente emocional en la lucha contra el DI. La implementación de técnicas de NLP fue importante en este estudio, ya que permitió la clasificación automatizada y el análisis de grandes volúmenes de datos. Este enfoque metodológico no solo validó la precisión del modelo en la detección de trinos con DI,

sino que también demostró su capacidad para manejar y analizar corpus diversos, lo cual es esencial en estudios de esta naturaleza.

Teóricamente, la investigación amplía el concepto de DI al explorar su manifestación en contextos de crisis de seguridad, como la ocurrida en Ecuador, un área menos investigada en comparación con otros escenarios como las elecciones, atentados terroristas o desastres naturales. Este trabajo proporciona nuevas perspectivas sobre cómo el DI se manifiesta en situaciones de crisis, donde la demanda urgente de información puede conducir a una mayor circulación de contenido con DI. Además, el desarrollo de un modelo basado en BERT para la detección automatizada del DI representa una contribución metodológica que ofrece un marco replicable para futuros estudios en entornos de crisis de seguridad. Este modelo facilita la identificación precisa de DI y permite un análisis en tiempo real de las dinámicas del DI.

No obstante, el estudio presenta limitaciones. El análisis se centró exclusivamente en X y en un período específico de la crisis en Ecuador, lo que podría limitar la generalización de los hallazgos a otras plataformas o a diferentes contextos temporales. Además, aunque el modelo de clasificación basado en BERT mostró un rendimiento sólido, su precisión podría mejorarse mediante la incorporación de un conjunto de datos más amplio y diversificado. Un corpus más extenso y variado permitiría afinar aún más el modelo al mejorar su capacidad para adaptarse a diferentes contextos y tipos de crisis, y aumentar la eficacia en la detección de DI en escenarios futuros.

Los resultados obtenidos plantean nuevas preguntas de investigación que podrían guiar estudios en este campo. Sería relevante explorar cómo se comporta el DI en otras plataformas digitales durante crisis de seguridad en

Latinoamérica y evaluar si existen patrones comunes o divergentes en la diseminación de DI y sus cargas emocionales. Además, investigar el impacto a largo plazo del DI en la percepción pública y la confianza en las instituciones podría ofrecer una comprensión más profunda de sus consecuencias en la sociedad, la polarización y la democracia.

La principal contribución de este artículo radica en establecer una relación entre las crisis de seguridad, como la experimentada en Ecuador, y la caracterización del DI en X, especialmente en su contenido emocional negativo, que puede incrementar la percepción de caos en la ciudadanía. Este estudio subraya la importancia de desarrollar herramientas automatizadas, como el modelo basado en BERT, para identificar y mitigar el DI, procurando que la información verificable prevalezca en crisis de seguridad y otras coyunturas similares.

## Referencias

- Abad, D. (2017). *Estadísticas de Facebook y Twitter en Colombia (2015)*. Social Insights.
- Abadía, A., Manfredi, L. C., & Sayago, J. T. (2023). Comunicación de crisis durante la pandemia del Covid-19 y su impacto en los sentimientos de la ciudadanía. *Opinión Pública*, 29(1), 199-225. <https://doi.org/10.1590/1807-01912023291199>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-236. <https://doi.org/10.1257/jep.31.2.211>
- Argüello Pazmiño, S., & Hurtado Arroba, E. (2024). Arenas públicas virtuales en Twitter y politización religiosa en Ecuador. *Estudios Sociológicos*, 42, 1-25. <http://dx.doi.org/10.24201/es.2024v42.e2329>
- Badrán, F. (2023). The signing of the peace agreement in Colombia. Old wine in new skins: Implications for national security and organized crime. *Trends in Organized Crime*. <https://doi.org/10.1007/s12117-023-09497-y>
- Barrios, M. M. (2019). When online commentary turns into violence: The role of Twitter in slander against journalists in Colombia. *Conflict & Communication Online*, 18(1). [https://regener-online.de/journalcco/2019\\_1/pdf/barrios-et-al2019.pdf](https://regener-online.de/journalcco/2019_1/pdf/barrios-et-al2019.pdf)
- BBC News Mundo. (2024, enero 17). Ecuador: Asesinan a fiscal que investigaba grandes casos de corrupción y el asalto armado a un canal de televisión. *BBC News Mundo*. <https://www.bbc.com/mundo/articles/c3gyg351mno>
- Blankstein, A., Williams, P., Elbaum, R., & Chuck, E. (2017, octubre 2). Las Vegas Shooting: 59 Killed and More Than 500 Hurt Near Mandalay Bay. *NBC News*. <https://www.nbcnews.com/storyline/las-vegas-shooting/las-vegas-police-investigating-shooting-mandalay-bay-n806461>
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., & Pérez, J. (2020). *Spanish pre-trained BERT model and evaluation data*. arXiv. <https://arxiv.org/abs/2308.02976>
- Cañizares, A. M., Alvarado, A., John, T., Rios, M., Suárez, K., & Stapleton, A. C. (2024, enero 10). Qué está pasando en Ecuador tras los hechos de violencia que sacuden el país. *CNN*. <https://cnnespanol.cnn.com/2024/01/10/ecuador-violencia-conflicto-armado-estado-excepcion-recap-trax/>
- Carrión Mena, F., & Silva, E. (2023). Ecuador en el concierto de la violencia de América Latina. *Ecuador Debate*, 119, 41-43.
- Castiblanco Briceño, M. E. (2024). Desórdenes informativos [base de datos]. *GitHub*. <https://github.com/manuelcastiblanco/desordenes-informativos>
- Chandra, R., & Saini, R. (2021). Biden vs Trump: Modeling US General Elections Using BERT Language Model. *IEEE Access*, 9, 128494-128505. <https://doi.org/10.1109/ACCESS.2021.3111035>
- Chavero, P., & Intriago, D. (2021). Las fake news como herramienta política durante la pandemia del COVID-19 en Ecuador. *Cuadernos del Centro de Estudios en Diseño y Comunicación. Ensayos*, (136), 19-35. <https://doi.org/ISSN-e-1853-3523,ISSN-1668-0227>
- Chen, L., Liu, Y., Chang, Y., Wang, X., & Luo, X. (2020). Public opinion analysis of novel coronavirus from online data. *Journal of Safety Science and Resilience*, 1(2), 120-127. <https://doi.org/10.1016/j.jnlsr.2020.08.002>
- Chen, X. K., Neo, L. S., Ang, A., Lee, J. Y., Ong, G., & Khader, M. (2022). The nexus between information disorder and terrorism: a mix of machine learning approach and content analysis on 39 terror attacks. *Dynamics of Asymmetric Conflict*, 15(3), 190-209. <https://doi.org/10.1080/17467586.2022.2055097>
- Cifuentes, C. F., & Pino Uribe, J. F. (2018). Conmigo o contra mí: Análisis de la concordancia y estrategias temáticas del Centro Democrático en Twitter. *Palabra Clave*, 21(3), 885-916. <https://doi.org/10.5294/pacla.2018.21.3.10>
- Cinelli, M., Cresci, S., Galeazzi, A., Quattrociocchi, W., & Tesconi, M. (2020). The limited reach of fake news on Twitter during 2019 European elections. *PLoS ONE*, 15(6), e0234689. <https://doi.org/10.1371/journal.pone.0234689>
- Corredoira, L. (2020). European regulatory responses to disinformation. Special attention to election campaigns. *Derecom*, (29), 5. <https://dialnet.unirioja.es/servlet/articulo?codigo=7596883>



- Crespo-Martínez, I., Garrido-Rubia, A., & Rojo-Martínez, J. M. (2022). El uso de las emociones en la comunicación político-electoral. *Revista Española De Ciencia Política*, (58), 175-201. <https://doi.org/10.21308/recp.58.06>
- Earley, M. (2021, septiembre 6). Auckland terror attack: Police warn of fake news circulating on social media. *Stuff*. <https://www.stuff.co.nz/national/300401074/auckland-terror-attack-police-warn-of-fake-news-circulating-on-social-media>
- El Economista. (2024, 15 de enero). Crisis en Ecuador: Agencia penitenciaria reporta el escape de 48 presos. <https://www.economista.com.mx/internacionales/Crisis-en-Ecuador-agencia-penitenciaria-reporta-el-escape-de-48-presos-20240115-0077.html>
- European Commission: Directorate-General for Communications Networks, Content and Technology. (2018). *A multi-dimensional approach to disinformation: report of the independent High level Group on fake news and online disinformation*. Publications Office. <https://data.europa.eu/doi/10.2759/739290>
- France 24. (2024, enero 15). Violencia en Ecuador dispara el número de homicidios de niños, según Unicef. <https://www.france24.com/es/minuto-a-minuto/20240115-violencia-en-ecuador-dispara-el-n%C3%BAmero-de-homicidios-de-ni%C3%B1os-seg%C3%BAn-unicef>
- Glatsky, G., & Cabrera, J. M. L. (2023, agosto 17). Ecuador lucha contra la peor ola de violencia generada por el narcotráfico. *The New York Times*. <https://www.nytimes.com/es/2023/08/17/espanol/ecuador-narcotrafico-elecciones.html>
- Gómez, L., Vallejo Mejía, M., Lombana-Bermúdez, A., & Pino Uribe, J. F. (2022). Encuadres en pugna por la paz en Twitter: El caso de las elecciones subnacionales del 2019. En F. Botero, J. F. Pino Uribe, B. Ortega, & L. Wills Otero (Eds.), *En configuración permanente: Partidos y elecciones nacionales y subnacionales en Colombia, 2018-2019* (pp. 255-283). Uniandes y Pontificia Universidad Javeriana.
- Guo, L., & Vargo, C. J. (2020). Fake news and emerging online media ecosystem: An integrated intermedia agenda-setting analysis of the 2016 US presidential election. *Communication Research*, 47(2), 178-200. <https://doi.org/10.1177/0093650218777177>
- Gutiérrez-Coba, L., & Rodríguez-Pérez, C. (2023). Estrategias de posverdad y desinformación en las elecciones presidenciales colombianas 2022. *Revista de Comunicación*, 22(2), 225-242.
- Harsin, J. (2015). Regimes of posttruth, postpolitics, and attention economy. *Communication, Culture & Critique*, 8(2), 327-333. <https://doi.org/10.1111/ccc.12097>
- Heller, M. (1980). *L'utopie au pouvoir. Histoire de l'URSS de 1917 à nos jours*. Calmann Levy.
- Innes, M. (2017). *Russian influence and interference measures following the 2017 UK terrorist attacks*. Cardiff University Crime and Security Research Institute.
- Johnson, J. (2018). The self-radicalization of white men: "Fake News" and the affective networking of Paranoia. *Communication, Culture and Critique*, 11(1), 100-115. <https://doi.org/10.1093/ccc/tcx014>
- Khyani, D., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2021). An Interpretation of Lemmatization and Stemming in Natural Language Processing. *Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology*, 22, 350-357.
- Kostakos, P., Nykanen, M., Martinviita, M., Pandya, A., & Oussalah, M. (2018). *Meta-terrorism: Identifying linguistic patterns in public discourse after an attack*. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, España, 2018, pp. 1079-1083. <https://doi.org/10.1109/ASONAM.2018.8508647>
- La Data Ec [@LaDataEc]. (2024, enero 16). *¿Se prohibirán los tatuajes para ser policías! Uno de los nuevos requisitos para aspirantes a policía nacional, será que no pueden tener tatuajes. Las autoridades aseguran que quieren un personal sin vínculos con las bandas criminales* [Posteo]. X <https://twitter.com/La-DataEc/status/1612468000000000000>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., & Zittrain, J. L. (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. *Science*, 359(6380), 1094-1096. <https://doi.org/10.1126/science.aao2998>
- Lombana-Bermúdez, A., Vallejo Mejía, M., Gómez Céspedes, L. M., & Pino Uribe, J. F. (2022). Cámaras de eco, desinformación y campañas de desprestigio en Colombia. Un estudio de Twitter y las elecciones locales de Medellín en 2019. *Política y Gobierno*, 29(1). <https://www.flacsoandes.edu.ec/sites/default/files/%25f/agora/files/FA-AGORA-2022-PINO.pdf>
- Manfredi, L. C., Abadía, A. A., & Rodríguez, J. L. (2024). Political engagement and aggressive use of social networks. Presidential campaigns in a highly polarized electoral scenario. En B. Di Fátima (Ed.), *Legal Challenges and Political Strategies in the Post-Truth Era—Online Hate Speech Trilogy* (Vol. II, pp. 67-90). LabCom Books & Editorial Universidad Icesi. <https://doi.org/10.18046/EUI/ohst.v1>
- Manfredi, L., & González, J. M. (2019). Comunicación y competencia en Twitter. Un análisis en las elecciones presidenciales Colombia 2018. *Revista Estudios Institucionales*, 6(11), 133-130. <https://doi.org/10.5944/eeii.vol.6.n.11.2019.25086>

- Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, 36(4), 102048. <https://doi.org/10.1016/j.jksuci.2024.102048>
- Marwick, A., & Lewis, R. (2017, mayo 15). Media manipulation and disinformation online. *Data & Society Research Institute*. <https://datasociety.net/library/media-manipulation-and-disinfo-online/>
- McClain, C., Anderson, M., Gelles-Watnick, R. (2024, junio 12). *How Americans Navigate Politics on TikTok, X, Facebook and Instagram*. Pew Research Center. <https://www.pewresearch.org/internet/2024/06/12/how-americans-navigate-politics-on-tiktok-x-facebook-and-instagram/>
- McIntyre, L. (2018). *Post-truth*. The MIT Press.
- Melo, J. D. M., Pino Uribe, J. F., Gómez, L. M., & Carvajal Pardo, A. (2023). Desinformación y democracia subnacional en Twitter: El caso de la campaña a la Alcaldía de Claudia López 2019 en Bogotá. *Democracias*, 11(11), 131-162. <https://doi.org/10.54887/27376192.74>
- Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., & Liang, X. (2018). *Doccano: Text Annotation Tool for Human* [software]. <https://github.com/doccano/doccano>
- Navarrete, I. (2024, enero 17). Abogado de presos demandará al Estado por intervenir en la cárcel de Esmeraldas. *El Diario Ecuador*. <https://www.eldiario.ec/actualidad/ecuador/abogado-de-presos-demandara-al-estado/>
- Navarro, I. (2021, noviembre 14). Rafael Correa tomó decisiones que beneficiaron al crimen organizado. *Milenio*. <https://www.milenio.com/internacional/rafael-correa-tomo-decisiones-beneficiaron-crimen-organizado>
- Noboa Gonzalez, M. F. (2024, enero 16). Ecuador: Cómo la 'isla de paz' de Sudamérica es ahora uno de los países más violentos del mundo. *The Conversation*. <http://theconversation.com/ecuador-como-la-isla-de-paz-de-sudamerica-es-ahora-uno-de-los-paises-mas-violentos-del-mundo-221197>
- Ortiz dos Santos, F. (2021). Myths and misconceptions on Covid-19: 'Congo Check' and 'Talato' verification experiences. *Frontiers in Communication*, 6. <https://doi.org/10.3389/fcomm.2021.627214>
- Página 12. (2024, enero 15). Los militares retomaron el control de varias cárceles de Ecuador | Más de 100 funcionarios y guardias penitenciarios fueron liberados. <https://www.pagina12.com.ar/704003-los-militares-retomaron-el-control-de-varias-carceles-de-ecu>
- Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A., & Martínez, M. V. (2024). *pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks* (No. arXiv:2106.09462). arXiv. <http://arxiv.org/abs/2106.09462>
- Piazza, J. A. (2021). Fake news: the effects of social media disinformation on domestic terrorism. *Dynamics of Asymmetric Conflict*, 15(1), 55-77. <https://doi.org/10.1080/17467586.2021.1895263>
- Pontón Cevallos, D., & Rivera Velez, F. (2024). Cinco perspectivas interpretativas sobre el incremento de la violencia en Ecuador. *Sociología y Política HOY*, (9), 139-167.
- Pozo Peralta, J. X. (2024). El Aumento de la Violencia en Ecuador: Análisis de las Variables Socioeconómicas y su Relación con la Tasa de Homicidios. *Revista de Cultura de Paz*, 7, 99-122.
- Primicias. (2024, enero 9). "Estamos vivos, fue un ataque sumamente violento", dice presentador de TC Televisión. <https://www.primicias.ec/noticias/seguridad/trabajadores-tc-television-canal-armados-guayaquil-ecuador>
- Ramos, I. (2013). Trayectorias de democratización y desdemocratización de la comunicación en Ecuador. *Íconos - Revista de Ciencias Sociales*, (46), 67. <https://doi.org/10.17141/iconos.46.2013.133>
- Razo, L. Á. H. (2021). Fake news, la contra-agenda mediática durante las elecciones federales en México 2021. *El Cotidiano*, 37(229), 93-103.
- Reuters. (2023, agosto 10). Explainer: Why has Ecuador become so violent? <https://www.reuters.com/world/americas/why-has-ecuador-become-so-violent-2023-08-10/>
- Rodríguez-Ferrándiz, R. (2023). An Overview of the Fake News Phenomenon: From Untruth-Driven to Post-Truth-Driven Approaches. *Media and Communication*, 11(2), 15-29. <https://doi.org/10.17645/mac.v11i2.6315>
- Salazar, A. (2024, 15 de enero). Asamblea dio paso a análisis del alza del IVA. *El Mercurio*. <https://elmercurio.com.ec/2024/01/15/asamblea-dio-paso-a-analisis-del-alza-del-iva/>
- Saló, C. C., & Pérez-Soler, S. (2022). Desorden informativo en la Unión Europea: construyendo una respuesta normativa. *Revista CIDOB d'Afers Internacionals*, (131), 141-161.
- Santana Chaves, J. M., & De Camargo Penteadó, C. L. (2023). Desórdenes informativos y polarización afectiva: En un análisis de las emociones en publicaciones con desinformación en tres conflictos políticos brasileños en Twitter. *Miguel Hernández Communication Journal*, 14, 67-88. <https://doi.org/10.21134/mhjournal.v14i.1811>
- Savransky, R. (2017, octubre 2). FBI: Las Vegas shooting has no connection with international terrorist group. *The Hill*. <https://thehill.com/homenews/news/353426-fbi-las-vegas-shooting-has-no-connection-with-international-terrorist-group/>
- Schwartz, J. (2017, octubre 2). Did some media play into ISIS's hands? *Politico*. <https://www.politico.com/story/2017/10/02/las-vegas-isis-shooting-media-243383>
- Siyech, M. S. (2019). *The pulwama attack's after effects: Fake news and rumours* (No. CO19034; RSIS Commentary). S. Rajaratnam School of International Studies.
- Son, G. H. W., & Rashid, E. I. A. (2022). *Classification of information disorder*. Khazanah Research Institute.

- Stremlau, N. (2022). Online speech and offline violence: Reflections on the current violence in Ethiopia. *Global Media Journal – German Edition*, 12(2). <https://doi.org/10.22032/dbt.55517>
- Taddeo, G., Frutos Torres, B. D., & Cruz Alvarado, M. (2022). Creadores y espectadores frente al desorden informativo online: efectos de la producción de contenidos digitales en competencias informativas. *Comunicar*, 72, 9-20. <https://doi.org/10.3916/C72-2022-01>
- Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news” A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153.
- Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, A., Stukal, D., & Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *SSRN*. <https://doi.org/10.2139/ssrn.3144139>
- Tufail, M., Song, L., Ali, S., Wahab, S., & Hassan, T. (2023). Does more immigration lead to more violent and property crimes? A case study of 30 selected OECD countries. *Economic Research-Ekonomika Istraživanja*, 36(1), 1867-1885. <https://doi.org/10.1080/1331677X.2022.2094437>
- Wan, X., Wang, R., Wang, M., Deng, J., Zhou, Z., Yi, X., Pan, J., & Du, Y. (2022). Online Public Opinion Mining for Large Cross-Regional Projects: Case Study of the South-to-North Water Diversion Project in China. *Journal of Management in Engineering*, 38(1), 05021011. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000970](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000970)
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Consejo de Europa. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>

**Disponibilidad de datos:** El conjunto de datos que apoya los resultados de este estudio se encuentra disponible en GitHub: <https://github.com/manuelcastiblan/desordenesinformativos>.

**Contribución de los autores (Taxonomía CRediT):** 1. Conceptualización; 2. Curación de datos; 3. Análisis formal; 4. Adquisición de fondos; 5. Investigación; 6. Metodología; 7. Administración de proyecto; 8. Recursos; 9. Software; 10. Supervisión; 11. Validación; 12. Visualización; 13. Redacción: borrador original; 14. Redacción: revisión y edición.

J. F. P. U. ha contribuido en 1, 5, 4, 10, 13, 14; M. E. C. B. en 1, 2, 3, 5, 6, 9, 10, 12, 13, 14; P. V. Z. C. en 2, 3, 5, 6, 11, 13, 14; J. D. M. V. en 1, 13.

Editor responsable: L. D.

## Apéndice A1

En la primera fase del estudio, el equipo de investigación inició un proceso de clasificación manual y análisis detallado de trinos para explorar el desorden informativo (DI) en el contexto de la crisis de seguridad en Ecuador. Este etiquetado inicial fue crucial, ya que permitió construir una base sólida sobre la cual aplicar técnicas avanzadas de procesamiento de lenguaje natural (NLP), haciendo posible clasificar grandes volúmenes de texto de forma precisa y eficiente. Partiendo de enfoques teóricos y metodológicos respaldados por estudios previos sobre desinformación y manipulación informativa, el equipo integró herramientas de NLP para refinar el análisis. A continuación, se detalla cada etapa de este proceso, describiendo las técnicas utilizadas y apoyándose en investigaciones clave que refuerzan la validez del método.

### Definición de categorías de desorden informativo

Para definir las categorías de DI, los investigadores se apoyaron en la tipología propuesta por Wardle y Derakhshan (2017), quienes distinguen tres formas de desorden informativo con base en la veracidad y la intención del contenido. Este marco conceptual ha sido ampliamente utilizado en estudios sobre desinformación (Allcott & Gentzkow, 2017; Lazer et al., 2018), y permite diferenciar entre distintos grados de manipulación y error en la información:

1. **Desinformación:** Se refiere al contenido falso o manipulado difundido con la intención explícita de engañar o causar daño, lo cual es frecuente en situaciones de alta tensión social (Piazza, 2021). Para identificar este tipo de trinos, el equipo buscó patrones de engaño deliberado, incluyendo el uso de lenguaje sensacionalista, la manipulación de imágenes y datos falsos que distorsionaran la realidad.
2. **Malinformación:** Consiste en la difusión de información verídica pero compartida con fines de perjudicar, revelando datos privados o sacando de contexto eventos sensibles. La malinformación puede intensificar la percepción pública de inseguridad y, en el caso ecuatoriano, se centró en el contexto de la crisis de seguridad. Siguiendo estudios como el de

Tandoc et al. (2018), el equipo evaluó la intención detrás del contenido, particularmente en los casos donde la exposición de datos sensibles apuntaba a fines perjudiciales.

3. **Misinformación:** Incluye errores y omisiones accidentales en el contenido compartido, tales como fechas incorrectas o estadísticas inexactas, sin intención de causar daño. En esta categoría, se verificaron los trinos contra fuentes oficiales, tomando en cuenta que, como sugieren Lazer et al. (2018), el impacto de la misinformación puede escalar en contextos de crisis y polarización, aunque se origine de errores involuntarios.

### Proceso de etiquetado y clasificación de trinos

Para el etiquetado y clasificación de trinos, el equipo empleó Doccano, una herramienta de etiquetado de texto ampliamente utilizada en estudios de NLP (Nakayama et al., 2018). El proceso se organizó en varias etapas, abordando tanto la preparación de los datos para el etiquetado como para su posterior análisis con el modelo:

#### 1. *Recolección y preparación de los datos para el etiquetado*

El equipo recolectó trinos publicados entre el 9 y el 17 de enero de 2024, un período marcado por un incremento significativo en la crisis de seguridad en Ecuador. Para capturar una diversidad de perspectivas y patrones de DI, el corpus se dividió en tres grupos de emisores: ciudadanos, medios de comunicación y asambleístas, siguiendo recomendaciones de estudios sobre crisis y polarización política (Gutiérrez-Coba & Rodríguez-Pérez, 2023).

#### 2. *Primera ronda de etiquetado manual*

En esta etapa inicial, cada investigador etiquetó un subconjunto de trinos aplicando las categorías de DI previamente definidas. Este etiquetado manual inicial proporcionó una línea base y permitió al equipo identificar patrones recurrentes de DI, una técnica efectiva en estudios similares (Chen et al., 2022). Además, la revisión cruzada entre los miembros del equipo ayudó a ajustar etiquetas y reducir sesgos de interpretación.

#### 3. *Revisión y validación de la clasificación*

Tras la primera ronda, el equipo realizó sesiones de revisión para resolver posibles ambigüedades y garantizar la coherencia en la clasificación. Esta validación intersubjetiva, frecuente en estudios de ciencias sociales y NLP (Wan et al., 2022), permitió afinar los criterios de cada categoría y adaptarlos al contexto específico ecuatoriano, caracterizado por un DI particularmente intenso en torno a la seguridad pública.

#### 4. *Filtrado de datos*

A continuación, el equipo filtró el corpus para mejorar la calidad de los datos finales. Se excluyeron trinos en otros idiomas, mensajes irrelevantes al tema de la crisis de seguridad y aquellos compuestos exclusivamente por emojis o ruido textual. Este filtrado es crucial en el procesamiento de texto, pues reduce sesgos en los resultados y mejora la precisión de los análisis subsiguientes (McIntyre, 2018).

#### 5. *Preparación de los datos para el modelo*

Finalmente, se realizó una limpieza exhaustiva del texto con el objetivo de preparar los datos para el análisis computacional. Este proceso incluyó la eliminación de menciones, *hashtags* y caracteres especiales, además de la lematización de palabras, para asegurar una mayor consistencia en el análisis posterior. La literatura, como el estudio de Khyani et al. (2021), ha demostrado que la lematización mejora la precisión en el análisis de sentimientos y facilita la clasificación de texto en proyectos de NLP.

### Implementación de modelos de NLP para clasificación automática

Para automatizar la clasificación de trinos, el equipo empleó el modelo BERT ajustado específicamente al contexto de la crisis en Ecuador. Esta elección se basó en la capacidad de BERT para identificar patrones semánticos y sintácticos complejos, lo que lo ha hecho altamente efectivo en estudios de clasificación de contenido y análisis de desinformación (Chandra & Saini,

2021). El modelo mostró un alto nivel de precisión en la detección de DI, con un desempeño superior al 92 % en las pruebas de validación.

Este enfoque metodológico no solo permitió procesar grandes volúmenes de datos con precisión, sino que también consolidó un sistema de clasificación de DI replicable para futuras investigaciones en otros contextos de crisis de seguridad y desinformación.

## Referencias

- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-236. <https://doi.org/10.1257/jep.31.2.211>
- Chandra, R., & Saini, R. (2021). Biden vs Trump: Modeling US General Elections Using BERT Language Model. *IEEE Access*, 9, 128494-128505. <https://doi.org/10.1109/ACCESS.2021.3111035>
- Chen, X. K., Neo, L. S., Ang, A., Lee, J. Y., Ong, G., & Khader, M. (2022). The nexus between information disorder and terrorism: a mix of machine learning approach and content analysis on 39 terror attacks. *Dynamics of Asymmetric Conflict*, 15(3), 190-209. <https://doi.org/10.1080/17467586.2022.2055097>
- Gutiérrez-Coba, L., & Rodríguez-Pérez, C. (2023). Estrategias de posverdad y desinformación en las elecciones presidenciales colombianas 2022. *Revista de Comunicación*, 22(2), 225-242.
- Khyani, D., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2021). An Interpretation of Lemmatization and Stemming in Natural Language Processing. *Shanghai Ligong Daxue Xuebao/ Journal of University of Shanghai for Science and Technology*, 22, 350-357.
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., & Zittrain, J. L. (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. *Science*, 359(6380), 1094-1096. <https://doi.org/10.1126/science.aao2998>
- McIntyre, L. (2018). *Post-truth*. The MIT Press.
- Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., & Liang, X. (2018). *Doccano: Text Annotation Tool for Human* [software]. <https://github.com/doccano/doccano>
- Piazza, J. A. (2021). Fake news: the effects of social media disinformation on domestic terrorism. *Dynamics of Asymmetric Conflict*, 15(1), 55-77. <https://doi.org/10.1080/17467586.2021.1895263>
- Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153.
- Wan, X., Wang, R., Wang, M., Deng, J., Zhou, Z., Yi, X., Pan, J., & Du, Y. (2022). Online Public Opinion Mining for Large Cross-Regional Projects: Case Study of the South-to-North Water Diversion Project in China. *Journal of Management in Engineering*, 38(1), 05021011. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000970](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000970)
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Consejo de Europa. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>