

GT 3.2. La transformación digital de los partidos políticos en las democracias occidentales: organización, campañas y participación en la era de las plataformas y la inteligencia artificial (2015-2025)

Incentivos algorítmicos, negatividad y toxicidad emocional en la comunicación electoral: estado de la cuestión e hipótesis de estudio

Bruno González-Cacheda,
Universidade de Vigo
bruno.gonzalez.cacheda@uvigo.gal

Fernando Martínez Arribas,
Universidade de Vigo
fernando.martinez@uvigo.gal

Resumen: La comunicación política canalizada a través de plataformas digitales y redes sociales, moldeada por algoritmos y mensajes personalizados, podría estar modificando el comportamiento de los actores políticos. En este sentido, algunos estudios muestran un incremento de los mensajes tóxicos y negativos de los partidos y líderes políticos durante las campañas electorales. Así, la indignación o el miedo producirían mejores resultados en términos de movilización y viralidad dentro de la estructura comunicativa moldeada por algoritmos y dirigida a audiencias segmentadas. De esta forma, los incentivos creados por este tipo de configuración comunicacional estarían moldeando el comportamiento de los partidos políticos y el tipo de mensaje utilizado en las campañas electorales, articulado en torno a la toxicidad y la negatividad. Por ello, analizaremos el estado de la cuestión a través de la revisión de los principales trabajos publicados con el objetivo de conocer las limitaciones, los consensos y las discusiones existentes en el ámbito académico. Además, la ejecución de las tareas señaladas nos permitirá avanzar en la delimitación de los principales elementos teórico-conceptuales. El texto concluirá con la presentación de las principales hipótesis tomando como referencia las elecciones de 23 de julio de 2023.

Palabras clave: Algoritmo, comunicación, redes, negatividad, toxicidad.

Nota biográfica de los autores

Bruno González Cacheda: Doctor en Gobernanza e Innovación. Licenciado en Ciencias Políticas y de la Administración y Diplomado en Gestión y Administración Pública. Profesor Ayudante Doctor de la *Universidade de Vigo*. Forma parte del grupo de innovación docente *Docencia Aberta_Open Teaching* y del grupo de investigación Observatorio de Gobernanza G³ de la *Universidade de Vigo*.

Fernando Martínez Arribas: Doctor en Derecho Público y Teoría del Estado. Graduado en Ciencia Política y de la Administración y Licenciado en Derecho. Profesor Permanente Laboral de la *Universidade de Vigo* y Profesor-Tutor en el Centro Asociado de la UNED de Lugo. Forma parte del grupo de innovación docente *Docencia Aberta_Open Teaching* y del grupo de investigación Observatorio de Gobernanza G³ de la *Universidade de Vigo*.

1. Introducción

La consolidación de las plataformas digitales y las redes sociales como espacio de comunicación política ha transformado las estrategias desarrolladas por los partidos políticos durante las campañas electorales. Así, sobre la base de la centralidad adquirida por los algoritmos, la economía de la atención y la competencia permanente por la visibilidad, los actores políticos adaptan sus estrategias y mensajes a la lógica de funcionamiento de los espacios digitales. En este contexto, la creciente presencia de contenidos asociados a las ideas de negatividad y la toxicidad ha despertado un notable interés académico debido a sus posibles implicaciones sobre la calidad del debate público y el funcionamiento de las democracias contemporáneas. La capacidad de las plataformas para amplificar determinados contenidos, favorecer la segmentación de audiencias y proporcionar información inmediata sobre el rendimiento de los mensajes ha generado un nuevo escenario comunicativo en el que la atención y la interacción se convierten en recursos estratégicos para los actores políticos. Como consecuencia, la comunicación electoral depende de forma creciente de la capacidad de producir contenidos competitivos dentro de las lógicas de difusión e interacción propias del ecosistema digital. En las siguientes líneas trataremos de indagar, a través de la revisión de la literatura publicada al respecto, hasta qué punto la configuración algorítmica de las plataformas y canales digitales podría modificar las estrategias comunicativas de los partidos a través del incremento en la utilización de recursos emocionales de carácter negativo o tóxico. Antes de abordar la literatura específica y los conceptos directamente relacionados, presentaremos el contexto en el que se circunscribe nuestro trabajo.

1.1. Contexto

Lo primero que debemos realizar al comenzar esta ponencia es situar el contexto en el que pretendemos hablar de los incentivos algorítmicos y la toxicidad emocional en la comunicación electoral.

Como señala Harari (2018), los “referéndum y las elecciones tienen siempre que ver con los sentimientos humanos, no con la racionalidad humana”. Si ese componente emocional se sitúa en el escenario digital, cabe plantearnos en qué medida la toma de decisión electoral (simplificando, el voto) puede estar influenciada por un algoritmo conocedor de nuestros gustos.

Pregunta que se realiza en un momento calificado como de “cuarta ola de la comunicación política digital” (García-Orosa, 2021). La primera, estaría datada en la década de los 90 del siglo XX, cuando los actores políticos comenzaron a interesarse por obtener presencia en internet creando webs; la segunda, entre 2004 y 2008, destacó por el desarrollo de redes sociales como *Facebook*, *Twitter* o *YouTube*, aunque los contenidos registraban todavía escasa adaptación a ellas; la tercera ola cobraría impulso con las elecciones de 2008 en Estados Unidos y la campaña electoral de Barack Obama, a través de un uso intensivo de las redes sociales, el diseño de mensajes personalizados, la microsegmentación y el protagonismo de los datos masivos (*big data*). En la actualidad, la cuarta ola a la que nos referíamos, cuyo origen se remonta al año 2016, coincidiendo con la campaña a favor del Brexit y los acontecimientos de *Cambridge Analytica*, se caracteriza porque: a) las plataformas digitales actúan como actores políticos que intervienen en todas las fases de comunicación; b) el uso intensivo de Inteligencia

Artificial (IA) y *big data*; c) la validación de la mentira como estrategia política (*fake news* y *deepfakes*); d) la combinación de elementos hiperlocales con los supranacionales; e) el determinismo acrítico tecnológico; f) la búsqueda de *engagement* con las audiencias y procesos de co-producción; y g) las tendencias con riesgos para la democracia: polarización de las opiniones, *astroturfing*, *eco chambers* y filtros burbuja (García-Orosa, 2021).

Por ello, aunque la cada vez mayor presencia de la IA en el proceso electoral puede tener efectos positivos (incremento de la participación derivada de la eventual implantación del voto electrónico remoto), los riesgos están a la orden del día (desinformación, ciberseguridad o sesgos). No puede obviarse que las injerencias digitales en los procesos electorales han llevado en algún caso a su anulación (Rumanía en el año 2024) dentro de las fronteras de la Unión Europea (UE). Por su parte, los sesgos y desinformación, con *deepfakes* generados por IA, pueden ser utilizados para confundir al electorado e incrementar la desconfianza hacia el proceso democrático. Basta recordar el autobús que en el referéndum del Brexit circuló por toda Gran Bretaña prometiendo destinar 350 millones de libras semanales a la sanidad pública en caso de salida de la UE. Frente a ello, empiezan a ponerse en marcha iniciativas a nivel global: el 21 de marzo de 2024, la Asamblea General de la Organización de las Naciones Unidas (ONU) adoptó la resolución 78/265 para promover sistemas de IA seguros y fiables que respeten los derechos humanos y contribuyan al desarrollo sostenible, seguida de la resolución 79/325, por la que se establece un Panel Científico Internacional Independiente sobre IA y el Diálogo Global sobre la Gobernanza de la IA.

En el ámbito interno español, no podemos olvidarnos del contexto en el que se desarrollaron las últimas elecciones generales realizadas hasta la fecha, el 23 de julio de 2023, convocadas anticipadamente por el presidente del Gobierno a raíz de los malos resultados obtenidos por su formación política, el Partido Socialista (PSOE), en las elecciones municipales y autonómicas de 28 de mayo de 2023.

La convocatoria, emanada del primer gobierno de coalición estatal habido desde la aprobación de la Constitución de 1978, se realizó en un ambiente de enorme polarización, en el que se sucedieron las referencias a la antigua organización terrorista ETA y, sobre todo, al *procés* catalán, que había desembocado en un referéndum anulado por el Tribunal Constitucional, la suspensión temporal del ejecutivo autonómico con la activación del artículo 155 de la Constitución y diversos procesos judiciales (incluyendo penas privativas de libertad) contra los principales dirigentes independentistas. Es en ese contexto donde el peso de Vox se hizo muy presente al condicionar los gobiernos autonómicos del Partido Popular (PP) formados a partir de los comicios de mayo, lo que a su vez contribuyó a una movilización de la izquierda, que acabó haciendo “bandera” de las políticas de igualdad y reconvirtió un insulto en un meme (“Perro Sanxe”) del que se acabó apropiando (Barragán-Romero, Caro-Castaño y Bellido-Pérez, 2024) como símbolo de autoidentificación. Polarización temática que también se trasladó a internet, pues los tres contenidos más presentes (Maldita-Democracy Reporting International, 2023) fueron “autonomía y nacionalismo” (18%), “derechos sociales” (16%) y “políticas de igualdad” (14%).

Estos comicios fueron los que mayor interés suscitaron entre el electorado (63,8%) desde 2008 (Garrido y Mora, 2024). De igual modo, el estudio realizado por Maldita-Democracy Reporting International (2023) monitorizando las publicaciones en *Twitter*, *Facebook*, *Instagram* y *YouTube* de un registro de más de 600 actores con vinculación política (incluyendo las cuentas oficiales de los partidos) generó un total de 142.000 publicaciones en los 30 días previos a las elecciones del 23 de julio, con un 32% de incremento respecto del período previo (que también había sido electoral), acumulándose el 15% los días de los debates electorales televisados (solo uno entre Pedro Sánchez y Alberto Núñez Feijóo). A la inversa, la digitalización hizo que fuera la campaña con el menor número de contactos directos entre la representación política y la ciudadanía de la serie histórica del Centro de Investigaciones Sociológicas (CIS), a la par que el promedio de los estudios demoscópicos se tradujo en una reducción de 3,5 puntos al PSOE respecto de los resultados que está formación realmente alcanzó (Garrido y Mora, 2024).

Precisamente, el análisis del papel de los incentivos algorítmicos y las emociones (sin olvidarnos de lo acontecido en los comicios del 23 de julio de 2023), así como de las hipótesis que podemos identificar, conformará el núcleo de este estudio.

1.2. Estado de la cuestión y relevancia del trabajo

En los últimos años, diversas investigaciones han ampliado el conocimiento relativo a las transformaciones introducidas por las plataformas digitales y las redes sociales en el ámbito de la comunicación política. Una parte de estos trabajos se ha centrado en la economía de la atención y en la importancia de las métricas de interacción para la adaptación comunicativa de los actores políticos a formatos y contenidos que generan mayor visibilidad (Franck, 2019; Ennser-Jedenastik et al., 2022; Michel y Gandon, 2024; Claesson, 2025). Desde otro punto de vista, un sector de la academia se ha focalizado en el estudio de cómo los mensajes negativos, los ataques y determinadas formas de comunicación tóxica tienden a alcanzar mayores niveles de difusión e interacción (Schöne et al., 2021; Fine y Hunt, 2023; Beknazar-Yuzbashev et al., 2025; Motte, 2026).

Sin embargo, la relación entre ambos planos no ha sido, hasta el momento, suficientemente explorada y desarrollada. Desde nuestro punto de vista, las evidencias publicadas permiten sostener que la configuración algorítmica de las plataformas y redes sociales generan incentivos que influyen en las características del contenido y el tono de los mensajes. No obstante, debemos señalar que no se puede afirmar de forma concluyente que los algoritmos sean la causa directa del incremento de la negatividad o la toxicidad. Por este motivo, resulta necesario avanzar en la articulación de un marco que conecte los incentivos propios del ecosistema digital con la proliferación y el incremento de estas formas de comunicación.

Otra de las cuestiones problemáticas se refiere a la delimitación conceptual de las principales ideas a través de las cuales articularemos nuestra investigación. Así, la negatividad, los ataques, las emociones negativas, la incivilidad y la toxicidad aparecen con frecuencia utilizados de forma confusa y escasamente justificada. Además, resulta frecuente observar cómo aparecen operacionalizados a partir de criterios diferentes, cuestión que dificulta la comparación entre resultados. Consideremos pues, que resulta imprescindible y pertinente la concreción de las principales dimensiones y propiedades

de las ideas de negatividad y toxicidad como conceptos relacionados, pero analíticamente diferenciados.

Para finalizar, entendemos necesario observar brevemente las investigaciones relacionadas con nuestra unidad de análisis y ámbito espacio-temporal. En este sentido, durante los últimos años han sido publicadas una serie de investigaciones relacionadas con nuestro objeto de estudio. Estos trabajos han analizado, entre otras cuestiones, el papel de las emociones en el comportamiento electoral durante las elecciones generales de 2023 (Jaráiz y García, 2024), las estrategias comunicativas desarrolladas por los liderazgos políticos en distintas plataformas digitales (Moreno-Cabanillas et al., 2026), la circulación de la desinformación y los discursos de odio durante la campaña electoral (Maldita-Democracy Reporting International, 2023) o la evolución de la toxicidad en la comunicación de los principales partidos políticos españoles entre 2015 y 2023 (Amores et al., 2025). Resulta evidente, que los estudios señalados han contribuido a mejorar la comprensión de los fenómenos descritos.

Sin embargo, observamos la carencia de una aproximación que articule de forma conjunta los incentivos generados por la configuración algorítmica de las redes sociales y la adaptación estratégica de los partidos a esas señales a través de la proliferación de mensajes cargados de negatividad y toxicidad. A través de la integración propuesta trataremos de comprender mejor la relación entre el entorno digital, el comportamiento de los partidos políticos en campaña y la tipología del mensaje emitido.

1.3. Objetivos

El objetivo general del presente trabajo consiste en analizar la posible relación entre los incentivos de interacción asociados a la configuración algorítmica de las redes sociales y la adaptación de las estrategias comunicativas de los partidos políticos a través de la utilización de mensajes negativos y tóxicos durante las campañas electorales

Para alcanzar este objetivo general se plantean los siguientes objetivos específicos:

- a) Revisar y analizar diversas publicaciones relativas a la economía de la atención, a la interacción digital y a la adaptación de la comunicación y los mensajes de los partidos políticos al entorno digital.
- b) Examinar la evidencia disponible relacionada con la difusión y el impacto comunicativo de los mensajes negativos y tóxicos en las redes sociales.
- c) Delimitar y concretar las dimensiones y propiedades conceptuales de negatividad y toxicidad.
- d) Identificar los factores políticos y contextuales que podrían favorecer la utilización de estas modalidades comunicativas.
- e) Formular hipótesis de investigación orientadas a un futuro análisis empírico de la comunicación desarrollada por el PP y el PSOE (en ese momento los únicos partidos con opciones reales de presidir el ejecutivo) durante la campaña de las elecciones generales de 2023.

2. Marco teórico

2.1. Los algoritmos y la economía de la atención

Según lo dispuesto por el Diccionario de la Real Academia de la Lengua Española, puede definirse algoritmo como “un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema”.

El concepto de algoritmo en redes sociales se refiere a una agrupación sofisticada de sistemas cuya misión primordial es incidir en el contenido, prioridad y frecuencia que se visualiza con el fin de maximizar el *engagement*. Es un hecho que los algoritmos analizan las búsquedas, gustos o historiales en internet, creando perfiles específicos sobre cada persona y condicionando la circulación de la información. Así, se puede hablar de la noción de “burbujas de filtro” (Bruns, 2022), derivada de la priorización algorítmica de contenido afin y que, al limitar las perspectivas diversas, incrementa los elementos de corte polarizador. Y también hay que tener presente que, en la actual organización de la sociedad, una parte considerable del poder deriva del control de la información que circula en las redes (Castells, 2006), usando los partidos políticos algoritmos para adaptar sus mensajes de forma masiva (*big data*), lo que cambia su modo de interacción con el electorado e incide en la forma en que se accede a los contenidos políticos y se elaboran las narrativas dentro del espacio digital (Estrada, 2026). Los procesos electorales se desarrollan cada vez más en espacios digitales, con mensajes rápidos, contenidos breves y narrativas visuales, configurándose nuevas dinámicas de construcción de la opinión pública en las que la visibilidad de los contenidos no depende únicamente de su valor informativo, sino de su capacidad para adaptarse a las lógicas algorítmicas (Berrocal-Gonzalo, Waisbord y Gómez-García, 2023). De esta forma, puede hablarse de una cibercultura en la que la información circula de forma rápida y descentralizada y prevalecen los impactos emocionales (Lévy, 2007), facilitando las redes la segmentación de las audiencias y la personalización de los mensajes, lo que a su vez incrementa la polarización política (Estrada, 2026).

Así es como podemos establecer un vínculo entre el algoritmo y la economía de la atención esbozada por Franck (2019), entendiendo la atención humana como un recurso limitado y valioso que los medios de comunicación compiten por captar. La fama o celebridad funciona como una bolsa de valores (medida en cifras de circulación y visualización, *likes*, etc.) que produce desigualdades cada vez mayores con unas tecnologías digitales que permiten monetizar la atención de forma mucho más eficaz que otros medios como la prensa. En una línea similar se pronuncian Michel y Gandon (2024), apuntando la necesidad de formular reglas estrictas para proteger la autonomía humana y la democracia de la manipulación en línea y la polarización. Por el contrario, otros sectores doctrinales señalan que hay que modular la importancia real de la desinformación (Metzler y García, 2024) estimando que representa una pequeña proporción del consumo de noticias digitales y no modifica fácilmente el comportamiento político en el voto (citando a Bail et al., 2020 y Guess et al., 2020); no obstante, admiten que los algoritmos podrían fomentar el discurso de odio, las noticias falsas y la polarización.

En esta línea, creemos necesario mencionar aportaciones de la literatura científica que buscan frenar las “burbujas de filtro”, obligando legalmente a las plataformas a exponer a sus usuarios a un porcentaje mínimo de contenidos con perspectivas ideológicas

opuestas y estableciendo medidas regulatorias que limiten la capacidad tecnológica de microsegmentación. Este sería el planteamiento de Acemoglu et al. (2025), impulsando programas públicos para identificar contenidos falsos y frenar la desinformación, con la finalidad de reducir la polarización del electorado y llevar a los partidos políticos hacia posiciones más moderadas.

En este contexto es donde podemos acudir de nuevo al plano interno, señalando el Proyecto de Ley Orgánica para el buen uso y la gobernanza de la Inteligencia Artificial, aprobado por el Consejo de Ministros el 26 de mayo de 2026, que prevé un mapeo de riesgos a partir del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de Inteligencia Artificial (RIA), y explicita los principios de supervisión humana, transparencia y protección de derechos, reforzando la transparencia algorítmica. No obstante, la fragmentación parlamentaria y el momento en el que se encuentra la actual legislatura no favorecen su aprobación en el corto plazo.

2.2. Incentivos emocionales y comunicación política de los partidos

De modo similar a lo que acabamos de señalar, también creemos posible establecer (al menos de forma preliminar) una relación entre algoritmo y el voto emocional del electorado.

Resulta innegable que las emociones ocupan un papel importante en la toma de decisiones políticas, incluso entre aquellas personas informadas, con conciencia política y que actúan habitualmente en base a criterios racionales (Lee y Choi, 2018). E igualmente lo es que el algoritmo tiene capacidad para incidir en esas emociones, pues el miedo o el conflicto obtienen la mayor cantidad de clics y quienes albergan estos sentimientos consumen más tiempo en las redes (Michel y Gandon, 2024). Estos autores reflejan cómo la burbuja de filtro se convierte en una burbuja emocional, en la que se conectan emociones preferentemente negativas, sesgos cognitivos (tendencia impulsiva a la indignación) y algoritmos, acentuada con mensajes cada vez más breves y simplistas (el lenguaje del emoticono) y por ciertas funcionalidades de las propias plataformas, facilitando algunas acciones (visibilidad de la opción “Me gusta”) a la vez que se dificultan otras (véase la denuncia de publicaciones, como señalan Michel y Gandon citando a Board, 2022). A esta posición podrían añadirse algunos estudios que muestran una vinculación entre la viralidad de las noticias falsas y la ira que provocan (Chuai y Zhao, 2022) y la contribución de ciertos algoritmos a favorecer información falsa que transmite ideas divisivas y emociones negativas, impulsando un extremismo desestabilizador de las democracias (Fortunato y Pecoraro, 2022).

Es cierto que existen sectores científicos que apuntan a la ausencia de evidencias que acrediten que los algoritmos causen burbujas de filtro y estas, a su vez, polarización (Metzler y García, 2024). Para ello mencionan un estudio realizado en 2014 con datos de *Facebook* (citando a Bakshy et al., 2015) que otorga mayor relevancia a las publicaciones realizadas que al algoritmo, apuntando respecto del de *YouTube* que solo 1 de cada 100.000 que empezaron visualizando contenido moderado transitó después al de extrema derecha (citando a Ribeiro et al., 2021). En sentido diferente, Allcott et al. (2020) constatan que en Estados Unidos la polarización se ha reducido cuando hay un menor uso de redes como *Facebook*. Ahora bien, la polarización y desinformación va más allá del

algoritmo o del empleo de redes sociales, no debiendo perderse de vista la incidencia del factor “analógico”, pues en su desarrollo puede tener un papel preponderante la inseguridad económica o una reacción adversa al cambio cultural progresista (Schäfer, 2022); aspectos que, en cualquier caso, logran un efecto multiplicador con internet.

Transitando de lo general a lo concreto, debemos mencionar el componente emocional de las concretas elecciones objeto de estudio (las generales de 2023), y su relación con los algoritmos presentes en internet y redes.

Es verdad que hay algún estudio sobre la variable emocional en dicho proceso electoral (Jaráiz y García, 2024), que destaca la naturaleza positiva de las emociones en el electorado de Sumar, la ansiedad hacia Yolanda Díaz en el votante de Vox, o el resentimiento hacia esta última formación como factor explicativo de voto al PSOE. Como ya hemos adelantado, se han llevado a cabo análisis sobre los contenidos presentes en redes durante el período previo a dichos comicios, evidenciando un elevado porcentaje de publicaciones dirigidas contra la integridad del proceso electoral (supuestas incidencias en el voto por correo o en el CERA) o vinculadas a discursos de odio a través de desinformaciones xenófobas.

Llegados a este punto, consideramos pertinente detenernos en la relación existente entre la configuración algorítmica de las redes sociales, caracterizada por la primacía de la emocionalidad, y la adaptación de las estrategias comunicativas de los partidos al escenario descrito. A este respecto, debemos comentar los hallazgos de Claesson (2025). Aunque la autora no se centra en la comunicación de campaña, los resultados de su trabajo son relevantes en la medida en la que analiza el comportamiento ordinario en materia de comunicación de los miembros de la Asamblea Nacional francesa. Claesson (2025) muestra que, a pesar de la opacidad algorítmica, los integrantes de la cámara legislativa adecúan sus mensajes al contexto digital a través del análisis de su impacto. De esta forma, según la misma autora, la adaptación sería reactiva y competitiva a través de la imitación y la copia de formatos exitosos en términos de interacción. A pesar de que el estudio no analiza específicamente la presencia de negatividad o toxicidad, sí ofrece indicios relevantes sobre la incorporación estratégica de la confrontación y la controversia. De hecho, alguno de los actores analizado reconoce que determinados conflictos son planificados anticipadamente por su capacidad para generar contenidos susceptibles de circular y obtener visibilidad en las redes. Por tanto, el trabajo no permite afirmar que la lógica algorítmica determine directamente la producción de mensajes negativos o tóxicos, pero sí evidencia que las métricas de interacción generan incentivos para adaptar la comunicación política hacia formas más emocionales, espectaculares y, en ciertos casos, deliberadamente confrontativas.

En un sentido semejante, Ennsner-Jedenastik et al. (2022), a partir del análisis de la comunicación de los principales partidos políticos austriacos en *Facebook* durante la campaña para las elecciones parlamentarias de 2017, muestran que las métricas de interacción proporcionadas por la plataforma condicionan las estrategias comunicativas de los partidos. En concreto, observan que aquellos temas que previamente habían generado mayores niveles de *engagement* tienen una mayor probabilidad de volver a ser incorporados a la agenda comunicativa. Esta tendencia se produce de forma más intensa en el caso de los partidos más importantes.

Más recientemente, el trabajo de Kim et al. (2025) profundiza en el planteamiento expuesto. De este modo, la adaptación estratégica iría más allá de la selección de los temas tratados, afectando también al tono de los mensajes emitidos. Tras analizar más de 56.000 anuncios de candidaturas al Congreso de Estados Unidos difundidos en *Facebook* durante las elecciones de 2020, este estudio muestra que los mensajes dirigidos a la captación de donaciones incorporan un mayor nivel de emociones negativas y toxicidad que aquellos orientados a la captación de votos. Por lo tanto, las características del discurso responderían a los incentivos derivados de los objetivos comunicativos y de las características de la audiencia destinataria. Debemos señalar que, en conjunto, los tres trabajos analizados apuntan a la adaptación de los actores políticos respecto a qué y a cómo comunican sus mensajes en función de los incentivos generados por el ecosistema digital.

En este contexto, la creciente adaptación de las estrategias comunicativas a las dinámicas de las plataformas digitales y redes sociales ha situado a las emociones en el centro de la comunicación política contemporánea. En conexión con las ideas sugeridas en los trabajos expuestos anteriormente, los partidos políticos ajustarían los temas y las características de sus mensajes en función de los incentivos generados por las métricas de interacción. Por otra parte, y dada la centralidad actual de las emociones como recurso persuasivo orientado a captar la atención y favorecer la movilización y la interacción (Martínez-Crespo et al., 2022), consideramos plausible que una parte importante de esa adecuación se materialice mediante el uso de recursos comunicativos articulados alrededor de componentes de carácter emocional. Así, en un entorno comunicativo caracterizado por la competencia permanente por la visibilidad y la atención, el mensaje de naturaleza emocional adquiere una relevancia creciente como mecanismo para maximizar el impacto y la interacción de los contenidos difundidos por los actores políticos.

En todo caso, debemos señalar que la dimensión emocional de la comunicación política no constituye una categoría homogénea. Como señalan Gerstlé y Nai (2019), las campañas electorales combinan diferentes estrategias comunicativas. Entre ellas, destaca el recurso a emociones como la esperanza, el entusiasmo, el miedo o la ansiedad, así como el empleo de campañas negativas basadas en el ataque al adversario. Partiendo de una visión plural de la naturaleza de los recursos emocionales en comunicación política, centraremos nuestra atención en dos dimensiones que, pese a encontrarse estrechamente relacionadas, responden a lógicas analíticas diferentes y han adquirido una relevancia creciente en el ecosistema digital: la negatividad y la toxicidad. Las próximas líneas estarán orientadas a la precisión conceptual de los rasgos que configuran ambas ideas.

2.3. Delimitación conceptual de la negatividad y la toxicidad

Como hemos subrayado en el anterior apartado, Fine y Hunt (2023) sostienen que los mensajes emitidos por los partidos y las élites políticas presentan una mayor potencia de difusión cuando incorporan elementos de lenguaje emocional intenso. En particular, señalan un mayor nivel de propagación a través de las redes sociales cuando se registran elementos vinculados a la negatividad emocional. Estos patrones de difusión generarían incentivos para que los actores políticos incrementen el recurso a este tipo de mensajes en sus estrategias de comunicación digital. En la misma línea, Schöne et al. (2021)

muestran que los mensajes con un mayor contenido negativo tienden a difundirse más ampliamente en las redes sociales, incluso cuando el contexto político es predominantemente positivo, reforzando así los incentivos para recurrir a este tipo de comunicación.

Sin embargo, bajo nuestro punto de vista, el creciente interés por el estudio de estos componentes comunicativos de tipo emocional no ha ido acompañado de un desarrollo y concreción conceptual equivalente. Así, aunque ambos fenómenos han sido ampliamente analizados desde perspectivas relacionadas con su difusión, sus efectos o su medición, su delimitación conceptual continúa siendo imprecisa (Lengauer et al., 2012; Soroka y McAdams, 2015; Fine y Hunt, 2023). Esta limitación se puede observar en la superposición y falta de claridad de nociones y dimensiones como negatividad, ataque, emociones negativas, incivilidad o toxicidad. La falta de consenso en la operacionalización de negatividad y toxicidad da lugar a una dificultosa comparación entre los resultados de los estudios relativos a este fenómeno. A modo de ejemplo, evidenciamos como Fine y Hunt (2023) utilizan la negatividad como un concepto amplio que engloba ataques, sentimientos y emociones negativas. Por el contrario, otros sectores la identifican con el tono emocional del mensaje (Schöne et al., 2021) o con el sesgo cognitivo asociado al procesamiento de información de carácter negativa (Soroka y McAdams, 2015). En los tres casos, observamos como elemento común la ausencia de delimitación respecto a la idea de toxicidad.

Tabla 1

Dimensiones y propiedades conceptuales de la negatividad y la toxicidad

Dimensión	Negatividad	Toxicidad
Destinatario	Actores sociopolíticos, individuos, políticas y programas e instituciones	Actores sociopolíticos e individuos
Función	Evaluar y/o criticar	Ofender, intimidar y/o deshumanizar
Tono	Desfavorable	Ofensivo y/o agresivo
Recursos empleados	Crítica, desacuerdo y/o confrontación	Insulto, amenazas, desprecio y/u odio

Nota. Elaboración propia

Una de las tentativas más elaboradas orientadas a la precisión conceptual de negatividad es la desarrollada por Lengauer et al. (2012), constatando la ausencia de una definición homogénea de esta idea. Para superar esta limitación, proponen entender la negatividad como una representación desfavorable de la política susceptible de manifestarse a través de diferentes dimensiones. Concretamente, sugieren que puede manifestarse de distintas formas: a) como una valoración desfavorable del tratamiento general de la información política; o b) como una crítica dirigida a actores políticos concretos. Partiendo de esta dicotomía, desarrollan una clasificación destinada a mejorar la comparabilidad de la investigación empírica.

Sin embargo, buena parte de la literatura posterior continúa utilizando el concepto de negatividad desde perspectivas diversas. Mientras Gerstlé y Nai (2019) la relacionan con los ataques dirigidos contra los adversarios políticos, Fine y Hunt (2023) la

operacionalizan mediante una combinación de ataques y sentimientos y emociones negativos. Por su parte, Schöne et al. (2021) equiparan la negatividad al tono emocional negativo del mensaje obtenido mediante técnicas de análisis de sentimiento. Finalmente, Soroka y McAdams (2015) centran su atención en la capacidad de la información negativa para captar la atención de los individuos sin precisar una definición explícita del propio concepto de negatividad.

Con el objetivo de superar los problemas conceptuales señalados, definiremos negatividad como la expresión de una valoración desfavorable sobre actores políticos, partidos, instituciones, políticas públicas o acontecimientos de naturaleza política. Esta valoración puede adoptar distintas formas, como la crítica, el desacuerdo, la confrontación, la acusación o el ataque (ver Tabla 1). Sin embargo, la presencia de estos elementos no implica necesariamente el uso de expresiones ofensivas o incívicas. De este modo, y bajo nuestro punto de vista, la negatividad constituye una forma de comunicación caracterizada por la emisión de valoraciones desfavorables sobre actores, instituciones, políticas o acontecimientos políticos.

Respecto a la noción de toxicidad, de forma similar a lo expuesto en las anteriores líneas, observamos la ausencia de una definición suficientemente consensuada. A pesar de que la literatura coincide en asociar toxicidad con formas de comunicación socialmente dañinas, los indicadores utilizados para medir las dimensiones que configuran este concepto presentan diferencias significativas entre estudios. Una de las propuestas más elaboradas es la de Pradel et al. (2024), quienes definen la toxicidad como un tipo de discurso perjudicial caracterizado por comportamientos comunicativos que vulneran las normas básicas de convivencia. Desde esta perspectiva, distinguen tres manifestaciones principales: la incivilidad, la intolerancia y las amenazas violentas. Esta aproximación sitúa el centro de la cuestión en los recursos discursivos empleados.

De modo semejante, diversas investigaciones coinciden en asociar la idea de toxicidad con el empleo de recursos discursivos ofensivos. Para ello utilizan una serie de indicadores coincidentes orientados a medir diversos elementos asociados a discursos de odio, acoso, lenguaje ofensivo, insultos, amenazas o ataques identitarios (Amores et al., 2025; Beknazar-Yuzbashev et al., 2025). En cambio, otros trabajos adoptan definiciones de carácter más restrictivo, centrando su atención en la utilización de insultos y expresiones hostiles dirigidas contra adversarios políticos (Motte, 2026).

A partir de la revisión realizada, definiremos toxicidad como una modalidad de comunicación caracterizada por el empleo de recursos discursivos ofensivos, degradantes o intimidatorios. Esta puede expresarse mediante insultos, amenazas, discursos de odio, descalificaciones o ataques identitarios, entre otras manifestaciones incompatibles con las normas básicas de respeto propias del debate democrático. En definitiva, tal y como se puede ver en la Tabla 1, distinguiremos negatividad y toxicidad como dos conceptos relacionados, pero analíticamente diferentes. Como hemos señalado, la negatividad implica una valoración desfavorable de la realidad política mientras que la toxicidad hace referencia al empleo de recursos discursivos ofensivos, degradantes o intimidatorios. O dicho de otro modo, la primera de las ideas vendría definida por el contenido de la valoración política expresada, mientras que la segunda aparecería asociada a la forma en la que se comunica dicha valoración (ver Tabla 1).

Para finalizar este apartado, más allá de su delimitación conceptual, debemos señalar que la utilización de mensajes negativos y tóxicos se encuentra vinculada a incentivos estratégicos derivados del entorno digital y del contexto político. Como ya hemos indicado, la lógica de funcionamiento de las plataformas sociales favorece aquellos contenidos capaces de generar mayores niveles de atención e interacción, incentivando el empleo de estrategias comunicativas basadas en la confrontación y la negatividad (Schöne et al., 2021; Fine y Hunt, 2023; Claesson, 2025; Milli et al., 2025). Además, estas características adquieren especial relevancia durante las campañas electorales, caracterizadas por un aumento de la competencia entre partidos y una intensificación del conflicto político. En este contexto, las redes sociales se convierten en un espacio central de movilización y disputa electoral (Moreno-Cabanillas et al., 2026).

Por otra parte, diversos trabajos han identificado factores contextuales asociados a la utilización de mensajes negativos y tóxicos en la comunicación política digital. Así, estas estrategias aparecen con mayor frecuencia vinculadas a partidos de derecha radical. Sin embargo, la evidencia no permite establecer una relación exclusiva ni uniforme con un determinado espacio ideológico (Daniel et al., 2024; Amores et al., 2025; Törnberg y Chueri, 2025) ya que su empleo depende del contexto competitivo en el que se desarrolla la comunicación política. Por lo tanto, la adopción de mensajes negativos y tóxicos no dependería únicamente de la orientación ideológica de los partidos, sino también de los incentivos derivados de la competencia política. La presencia de partidos de derecha radical que recurren intensamente a estrategias comunicativas basadas en la confrontación podría ejercer presión sobre otras formaciones del espacio electoral próximo para adoptar estilos discursivos más agresivos con el fin de evitar pérdidas de visibilidad o apoyo electoral. Esta dinámica se vería reforzada en las redes sociales, donde los mensajes negativos y tóxicos generan mayores niveles de interacción y difusión (Motte, 2026). En el caso español, la elevada toxicidad observada en la comunicación de Vox (Amores et al., 2025) y el contexto de intensa competencia entre PP y Vox durante las elecciones generales de 2023 permitirían plantear que dicha presión competitiva podría haber favorecido la adopción de un tono más negativo y tóxico por parte del Partido Popular.

2.4. Hipótesis de estudio

H1.- Durante la campaña de las elecciones generales españolas de 2023, la comunicación de los principales partidos políticos presentó un predominio de mensajes negativos sobre mensajes positivos o neutros.

H2.- La toxicidad aparece con menor frecuencia que la negatividad.

H3.- El PP presenta mayores niveles de negatividad y toxicidad que el PSOE, como consecuencia de la presión competitiva ejercida por Vox dentro del mismo espacio electoral.

Referencias bibliográficas

- Acemoglu, D., Ozdaglar, A., & Siderius, J. (2025). AI and social media: A political economy perspective. National Bureau of Economic Research, w33892. <https://ssrn.com/abstract=5296536>
- Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3), 629-676. <https://doi.org/10.1257/aer.20190658>
- Amores, J. J., González-Baquero, W., & Oller-Alonso, M. (2025). Toxicity levels in Spanish political communication on Twitter/X: A comparative analysis of major political parties, 2015–2023. *Frontiers in Political Science*, 7, 1627474. <https://doi.org/10.3389/fpos.2025.1627474>
- Barragán-Romero, A. I., Caro-Castaño, L., & Bellido-Pérez, E. (2024). La apropiación partidista del meme: fandom y propaganda en las elecciones generales españolas de 2023. *Revista Latina De Comunicación Social*, (83), 1–25. <https://doi.org/10.4185/rlds-2025-2304>
- Beknazar-Yuzbashev, G., Jiménez-Durán, R., McCrosky, J., & Stalinski, M. (2025). *Toxic content and user engagement on social media: Evidence from a field experiment* (No. 11644). CESifo Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5130929
- Berrocal-Gonzalo, S., Waisbord, S., & Gómez-García, S. (2023). Polarización política y medios de comunicación, su impacto en la democracia y en la sociedad. *Profesional de la Información*, 32(6). <https://doi.org/10.3145/epi.2023.nov.22>
- Bruns, A. (2022). El Filtro burbuja (*The Filter Bubble*). *Revista Latinoamericana de Economía y Sociedad Digital (RLESD)*. <https://eprints.qut.edu.au/235421/>
- Castells, M. (2006). *La sociedad red: una visión global*. Alianza Editorial.
- Chuai, Y., & Zhao, J. (2022). Anger can make fake news viral online. *Frontiers in Physics*, 10. <https://doi.org/10.3389/fphy.2022.970174>
- Claesson, A. (2025). Influencers in Parliament? Platform Pressures and Normative Conflict in Political Social Media Adoption. *Social Media + Society*, 11(4). <https://doi.org/10.1177/20563051251395410>
- Consejo de Ministros (2026). *Proyecto de Ley Orgánica para el buen uso y la gobernanza de la Inteligencia Artificial*. Aprobado el 26 de mayo de 2026 (BOCG, Congreso de los Diputados, 12 de junio de 2026).
- Daniel, W. T., Frelin, E., Robert, M.-V., & Rowley-Abel, L. (2024). Individual drivers of toxicity in radical right-wing populist legislative campaigns. *West European Politics*, 48, 590-617. <https://doi.org/10.1080/01402382.2024.2311040>
- Ennser-Jedenastik, L., Gahn, C., Bodlos, A., & Haselmayer, M. (2022). Does social media enhance party responsiveness? How user engagement shapes parties' issue attention on Facebook. *Party Politics*, 28(3), 468-481. <https://doi.org/10.1177/1354068820985334>

- Estrada, P. J. (2026). *La influencia de los algoritmos en la polarización política (elecciones presidenciales 2025) en Instagram*. UPS Ecuador.
- Fine, J. A., & Hunt, M. F. (2023). Negativity and elite message diffusion on social media. *Political Behavior*, 45(3), 955-973. <https://doi.org/10.1007/s11109-021-09740-8>
- Fortunato, P., & Pecoraro, M. (2022). Social media, education, and the rise of populist Euroscepticism. *Humanities and Social Sciences Communications*, 9(301). <https://www.nature.com/articles/s41599-022-01317-y>
- Franck, G. (2019). The economy of attention. *Journal of Sociology*, 55(1), 8-19. <https://doi.org/10.1177/1440783318811778>
- García-Orosa, B. (2021). Disinformation, social media, bots, and astroturfing: the fourth wave of digital democracy. *Profesional de la Información*, 30(6). <https://doi.org/10.3145/epi.2021.nov.03>
- Garrido, A., & Mora, A. (2024). La campaña electoral de 2023. En E. Jaraiz, C. Ortega y P. Oñate (coords.), *Elecciones generales de 2023* (pp. 39-56). CIS.
- Gerstlé, J., & Nai, A. (2019). Negativity, emotionality and populist rhetoric in election campaigns worldwide, and their effects on media attention and electoral success. *European Journal of Communication*, 34(4), 410-444. <https://doi.org/10.1177/0267323119861875>
- Harari, Y. N. (2018). *21 lecciones para el siglo XXI*. Debate.
- Jaraiz, E., & García, M. (2024). El impacto de las emociones en el voto. En E. Jaraiz, C. Ortega y P. Oñate (coords.), *Elecciones generales de 2023* (pp. 159-176). CIS.
- Kim, S. Y. S., Zilinsky, J., & Brew, B. (2025). Donate to help us fight back: Political fundraising and toxic rhetoric online. *Party Politics*, 31(3), 549-561. <https://doi.org/10.1177/13540688241235901>
- Lee, J., & Choi, Y. (2018). Expanding affective intelligence theory through social viewing: Focusing on the South Korea's 2017 presidential election. *Computers in Human Behavior*, 83, 119-128. <https://doi.org/10.1016/j.chb.2018.01.026>
- Lengauer, G., Esser, F., & Berganza, R. (2012). Negativity in political news: A review of concepts, operationalizations and key findings. *Journalism*, 13(2), 179-202. <https://doi.org/10.1177/1464884911427800>
- Lévy, P. (2007). *Cibercultura*. Anthropos Editorial.
- Maldita-Democracy Reporting International (2023). *Desinformación, odio y la conversación online de los políticos durante las elecciones generales de 2023*. <https://maldita.es/malditobulo/20230808/elecciones-generales-2023-desinformacion-politicos-odio/>
- Martínez-Crespo, I., Garrido-Rubia, A. G., & 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>

Metzler, H., & García, D. (2024). Social drivers and algorithmic mechanisms on digital media. *Perspectives on Psychological Science*, 19(5), 735-748. <https://doi.org/10.1177/17456916231185057>

Michel, F., & Gandon, F. (2024). Pay Attention: A call to regulate the attention market and prevent algorithmic emotional governance. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), 971-983. <https://doi.org/10.1609/aies.v7i1.31696>

Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS Nexus*, 4(3), pgaf062. <https://doi.org/10.1093/pnasnexus/pgaf062>

Moreno-Cabanillas, A., Castellero-Ostio, E., & Serna-Ortega, Á. (2026). Electoral campaigns on social networks: Leaders' strategies in Spain's 2023 general election. *Revista de Comunicación de la SEECI*, 59, 1-23. <https://doi.org/10.15198/seeci.2026.59.e921>

Motte, E. (2026). *Insult Politics in the Age of Social Media*. <https://elliott-motte.github.io/#working-papers>

Organización de las Naciones Unidas (2024). *Resolución 78/265 aprobada por la Asamblea General el 21 de marzo de 2024. Aprovechar las oportunidades de sistemas seguros y fiables de inteligencia artificial para el desarrollo sostenible*.

Organización de las Naciones Unidas (2025). *Resolución 79/325 aprobada por la Asamblea General el 26 de agosto de 2025. Mandato y modalidades para el establecimiento y funcionamiento del Panel Científico Internacional Independiente sobre Inteligencia Artificial y del Diálogo Mundial sobre la Gobernanza de la Inteligencia Artificial*.

Pradel, F., Zilinsky, J., Kosmidis, S., & Theocharis, Y. (2024). Toxic speech and limited demand for content moderation on social media. *American Political Science Review*, 118(4), 1895-1912. <https://doi.org/10.1017/S000305542300134X>

Schäfer, A. (2022). Cultural backlash? How (not) to explain the rise of authoritarian populism. *British Journal of Political Science*, 52(4), 1977-1993. <https://doi.org/10.1017/S0007123421000363>

Schöne, J. P., Parkinson, B., & Goldenberg, A. (2021). Negativity spreads more than positivity on Twitter after both positive and negative political situations. *Affective Science*, 2(4), 379-390. <https://doi.org/10.1007/s42761-021-00057-7>

Soroka, S., & McAdams, S. (2015). News, politics, and negativity. *Political Communication*, 32(1), 1-22. <https://doi.org/10.1080/10584609.2014.881942>

Törnberg, P., & Chueri, J. (2025). Elite Political Discourse has Become More Toxic in Western Countries. *ArXiv*. <https://doi.org/10.48550/arXiv.2503.22411>

Unión Europea (2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial*.