




# Miedo (in)fundado al algoritmo: Las recomendaciones de YouTube y la polarización

(Un)founded fear towards the algorithm:  
YouTube recommendations and polarisation

 Dr. Javier García-Marín. Profesor Titular, Departamento de Ciencia Política y de la Administración, Universidad de Granada (España) (jgmarin@ugr.es) (<https://orcid.org/0000-0002-2766-0266>)

 Dr. Ignacio-Jesús Serrano-Contreras. Investigador Asociado, Grupo de Investigación SINAI, Universidad de Jaén (España) (ijserran@ujaen.es) (<https://orcid.org/0000-0002-2399-0647>)

## RESUMEN

Las redes sociales han instaurado una nueva forma de comunicarse y entender las relaciones sociales. A su vez, en lo que podría entenderse como un aspecto negativo, los algoritmos se han construido y desarrollado bajo el paraguas de un amplio abanico de conjeturas y diferentes posiciones al respecto de su capacidad para dirigir y orquestrar la opinión pública. El presente trabajo aborda, desde los procesos de ingeniería inversa y de minado semántico, el análisis del sistema de recomendación de YouTube. De este modo, y, en primer lugar, reseñar un resultado clave, las temáticas analizadas de partida no tienden a extremarse. Seguidamente, y mediante el estudio de los temas seleccionados, los resultados no ofrecen una clara resolución de las hipótesis propuestas, ya que, como se ha mostrado en trabajos parecidos, los factores que dan forma al sistema de recomendación son variados y de muy diversa índole. De hecho, los resultados muestran cómo el contenido polarizante no es igual para todos los temas analizados, lo que puede indicar la existencia de moderadores –o acciones por parte de la compañía– que alteran la relación entre las variables. Con todo ello, trabajos como el presente abren la puerta a posteriores incursiones académicas en las que trazar sistematizaciones no lineales y con las que, tal vez, poder arrojar un sustento más neto y sustancial que permita despejar por completo parte de las dudas sobre el papel de los algoritmos y su papel en fenómenos sociales recientes.

## ABSTRACT

Social media have established a new way of communicating and understanding social relationships. At the same time, there are downsides, especially, their use of algorithms that have been built and developed under their umbrella and their potential to alter public opinion. This paper tries to analyse the YouTube recommendation system from the perspectives of reverse engineering and semantic mining. The first result is that, contrary to expectations, the issues do not tend to be extreme from the point of view of polarisation in all cases. Next, and through the study of the selected themes, the results do not offer a clear answer to the proposed hypotheses, since, as has been shown in similar works, the factors that shape the recommendation system are very diverse. In fact, results show that polarising content does not behave in the same way for all the topics analysed, which may indicate the existence of moderators –or corporate actions– that alter the relationship between the variables. Another contribution is the confirmation that we are dealing with non-linear, but potentially systematic, processes. Nevertheless, the present work opens the door to further academic research on the topic to clarify the unknowns about the role of these algorithms in our societies.

## PALABRAS CLAVE | KEYWORDS

Aprendizaje de máquina, YouTube, redes sociales, sistemas de recomendación, polarización, comunicación.  
Machine learning, YouTube, social media, recommendation system, polarisation, communication.

## 1. Introducción y estado de la cuestión

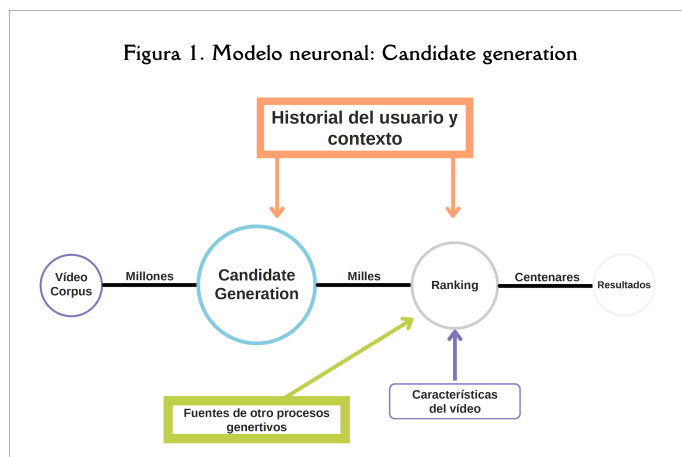
El auge de las TIC ha producido nuevos escenarios tanto para la interacción social como para el consumo de información y entretenimiento en las sociedades de finales del siglo XX, aunque con especial hincapié en las primeras décadas del presente siglo XXI. A los ya conocidos medios de masas, entre los que han destacado la prensa, la radio y la televisión, se fueron añadiendo otros métodos con los que difundir y acceder a mensajes de un modo masivo, lo que también generó nuevas simbiosis en donde los papeles de la cadena de comunicación tendieron a difuminarse e hibridarse (Berrocal-Gonzalo et al., 2014). Hablamos, en este caso, de las bondades y contribuciones que Internet ha incorporado para democratizar y nivelar los mecanismos de difusión de la información (Arias-Maldonado, 2016; Nielsen & Fletcher, 2020). Mientras que Internet y sus satélites iban creciendo —aupados por el desarrollo de la computación y los avances tecnológicos de procesamiento—, también lo hacían los múltiples enfoques sobre sus efectos. A este respecto, esa tela de araña (inter)conectada de la que hablaban diferentes autores (Berners-Lee, 2000), y con distintos epítetos (véanse McLuhan (1959), Habermas (1981) o Castells (2001) como un breve compendio de estas expresiones), ha engendrado primero las Web 1.0 y posteriormente las 2.0 (O'Reilly & Battelle, 2009), además de otras aún en vías de concreción (Latorre, 2022).

A raíz de los correspondientes avances, emanan variadas reflexiones al respecto de los efectos que pueden tener en los ciudadanos. Algunas de estas nuevas reflexiones ya señalaban cómo el auge de las crecientes Web 2.0 podrían convertirse en ejes de un cambio de paradigma, suponiendo a la postre un conjunto de retos y oportunidades tanto para las esferas políticas como para las democracias liberales del siglo XXI (Sunstein, 2007; Lilleker & Jackson, 2008; Chadwick, 2009; Howard, 2021; Messina, 2022). Pese a ello, aún no habían terminado de solidificar los grandes avances del mundo interconectado. El surgimiento de lo que ha pasado a denominarse redes o medios sociales propicia, sin embargo, una reconfiguración del desarrollo de las relaciones humanas (Wigand et al., 2010). Así, la incorporación al día a día de Facebook, Twitter, Instagram o YouTube ha supuesto una alteración para las sociedades contemporáneas, pudiendo reseñar de forma descriptiva, y entre otros factores a tener en cuenta, el albergar a usuarios de todos los rincones del planeta, además de contar entre sus ratios de alcance y difusión con más de un tercio de la población mundial. Inmersos ya en este periodo de tránsito, los focos ahora se centran en los efectos que se pueden dar en esta nueva deriva social y mediática conducida por las redes. A este respecto, parte de las prospecciones que se han llevado a cabo en la academia, pero también en otras esferas como las del periodismo o la política, han virado en torno a la cara interna de la interconexión. Hablamos, en este caso, del papel que pueden jugar los algoritmos, en especial sus arquitecturas y protocolos como mediadores del proceso de comunicación. Es ahora, y con la perspectiva que puede suponer esta expansión del factor algorítmico, cuando surgen cuestiones que pudieran entenderse aledañas a esta nueva deriva. En este contexto, emergen planteamientos que señalan a la posible concomitancia presente entre el papel de las redes sociales —aunque con especial énfasis en sus modelos computacionales— y la búsqueda de un entendimiento a sucesos sociales singulares desde una mayor presencia de polarización política en el debate público (Hernández et al., 2021) a situaciones complejas de definir como el Brexit o la victoria de Donald Trump en 2016.

Pese a las claras correlaciones que pudieran darse entre unos hechos y otros, lo cierto es que las investigaciones apuntan a diferentes lados, sin dejar nada por cerrado. Al menos así se establece en las dudas que presentan trabajos como los de Rasmussen y Petersen (2022), Bail (2021) o Barberá (2020) quienes señalan la multifactorialidad, incluso los que apuntan al plano analógico (Arceneaux & Johnson, 2010) como eje clave para poder alcanzar una respuesta. Por tanto, esta situación es compleja y sujeta a dinámicas ambivalentes. En base a todo ello, la presente investigación ahonda en parte de los fenómenos que hay tras YouTube y su capacidad de inundar las múltiples esferas de la escena mediática (Banaji, 2013). Uno de los factores clave, tal y como también señalan Yesilada y Lewandowsky (2022), se centra en la complejidad para alcanzar un conocimiento de su sistema. De este modo, y como ya apuntaron otros estudios como los de Luengo et al. (2021) o Serrano-Contreras et al. (2020), el presente trabajo tiene como objetivo señalar las derivas que el algoritmo puede generar. Además, busca incorporar al debate si ese modelo computacional supone algún rédito para considerar la aparición de fenómenos sociales como el de la polarización (Van-Bavel et al., 2021).

## 2. Material y métodos

Esta investigación propone realizar un análisis del algoritmo de YouTube desde múltiples perspectivas. A este respecto, bajo lo denominado como proceso de ingeniería inversa (Rekoff, 1985), así como mediante la implementación de técnicas de minado de texto y el uso de índices de medición semántica, se busca dar forma a un progresivo conocimiento de qué se halla bajo las arquitecturas computacionales implementadas por YouTube.



Nota. Modelo expuesto en Covington et al. (2016).

Para ello, buscamos dejar de lado algunas de las funciones principales que Alphabet, como dueño de YouTube, incorpora al entrenamiento, y posterior desarrollo, del algoritmo de su servidor de vídeos (véanse los trabajos de los propios empleados de Alphabet, Davidson et al. (2010) o Covington et al. (2016), así como la Figura 1, para una aproximación al modelo que emplea la plataforma para ofrecer los resultados al usuario). De este modo, la pretensión es tratar de parametrizar el comportamiento del modelo, por un lado, eludiendo el conjunto de datos pasivos y activos que en la red hacemos mientras navegamos (ej. ubicación; historial de búsqueda; datos personales...), ya que todas estas métricas sirven para la confección pormenorizada de nuestros supuestos intereses -fundamentales para nuestra experiencia de uso (Dimopoulos et al., 2013)-; mientras que, por otro lado, buscamos hacer una reseña del mismo. Estas experiencias son la base para supuestos indispensables a la hora de entender algunas de las nomenclaturas más usadas al hablar de algoritmos, como puede ser el filtro burbuja (Pariser, 2017), así como otro tipo de externalidades (Bishop, 2018).

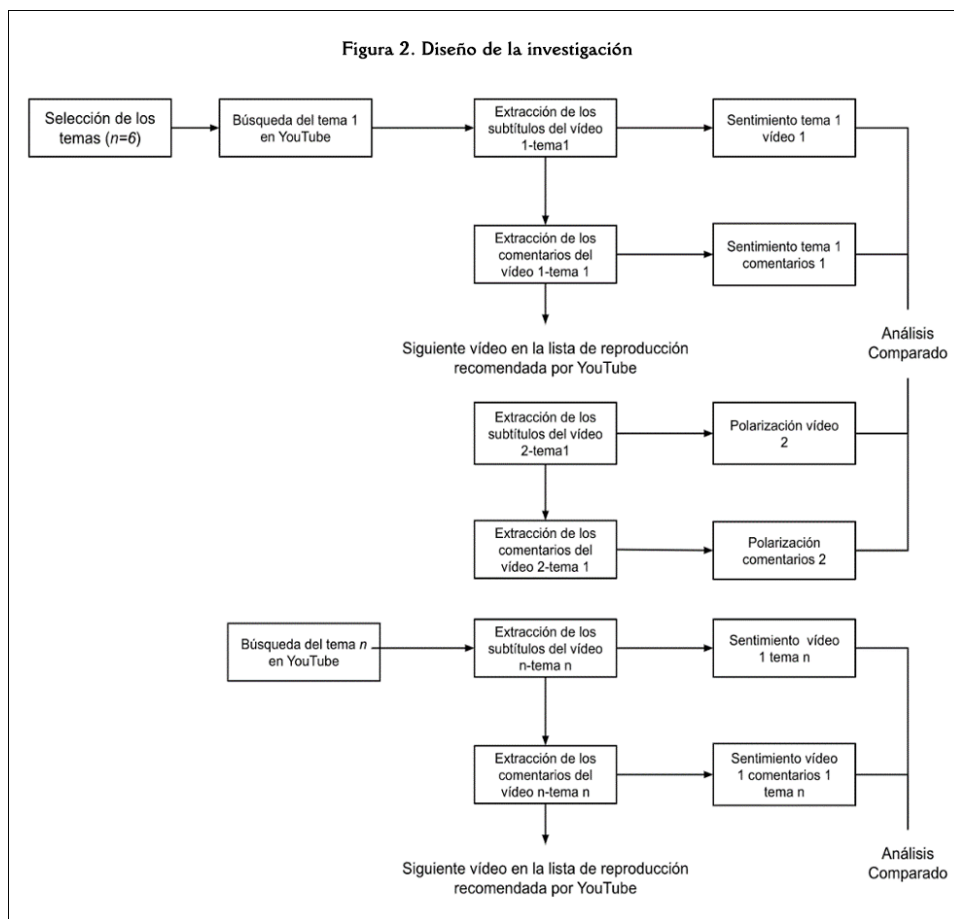
Dentro de este conjunto de efectos que la era algorítmica puede ocasionar se encontraría un factor muy manido recientemente, la idea de radicalización. Para dar sustento a estas ideas acerca del refuerzo de una postura, empleamos trabajos, investigaciones y ejemplos empíricos, que han mostrado cómo el algoritmo de la plataforma tendía a extremarse cada vez más (Tufekci, 2018; Alfano et al., 2021; Almagro & Villanueva, 2021; Chen et al., 2021). Por lo tanto, el comportamiento del algoritmo de recomendación de YouTube, al estar basado en los intereses mostrados por los usuarios, debería llevar a un mayor consumo de materiales relacionados. Es decir, que interpretará las búsquedas de los usuarios como intereses y, así, le intentará facilitar la labor de búsqueda mediante la recomendación de vídeos similares (lo que daría lugar a una burbuja de filtrado). Naturalmente, de eso se desprende que las recomendaciones podrían redundar en una mayor polarización al provocar una menor exposición a puntos de vista, o temas, diferentes. De ahí que la primera hipótesis que planteamos en la presente investigación sea:

- H1: Los vídeos recomendados por YouTube serán cada vez más extremos o polarizados.

Al mismo tiempo, dichas recomendaciones podrían tener el efecto de crear comunidades con intereses muy similares (lo que se puede denominar de clasificaciones homófilas) lo que, a su vez, podría estar relacionado con las cámaras de eco. Por eso, nuestra segunda hipótesis de investigación es:

- H2: Los comentarios a los vídeos recomendados por YouTube estarán igualmente polarizados.

La lógica de la investigación es que, según parte de la comunidad científica<sup>1</sup>, las recomendaciones del algoritmo de YouTube caminan hacia la extrapolación de los intereses planteados por las búsquedas de los usuarios, creando esa burbuja de filtrado y, al mismo tiempo, aumentando la polarización de los contenidos mostrados (al estar cada vez más centrados en un contenido o punto de vista muy específico). Los usuarios, así, deberían comportarse de forma similar.



La Figura 2 describe la investigación que hemos diseñado para la comprobación de las hipótesis. Primeramente, se ha seleccionado una cuenta existente de YouTube, perteneciente a uno de los investigadores, pero que nunca ha sido utilizada (podríamos decir que no hay metadatos iniciales vinculados a la cuenta, por lo que las búsquedas que iniciemos crearán los metadatos sobre nuestros intereses)<sup>2</sup>. Posteriormente, se ha procedido a la elección de temas sobre los que aplicar el diseño. Se ha intentado que los temas sean relevantes y que sirvan para discriminar algunos moderadores que pudieran afectar a la relación entre variables que pretendemos analizar. Así, los temas<sup>3</sup> son:

- Política nacional: búsqueda de «política nacional».
- Partidos políticos en los extremos ideológicos: búsquedas de «podemos» y «vox».
- Vegetarianismo: búsqueda de «comida vegetariana».
- Conflicto: búsqueda de «guerra de ucrania».
- Feminismo: Búsqueda de «feminismo».
- COVID: búsqueda de COVID.

Es decir, temas que difieren en la conflictividad y relacionados con diversos campos, desde los conflictos hasta el vegetarianismo. Todos son temas sensibles donde esperamos una cantidad significativa de comentarios (en los vídeos que lo permitan).

El paso siguiente fue iniciar la búsqueda del tema en YouTube, tanto en la cuenta propuesta como sin cuenta alguna, y el análisis del primer vídeo que nos recomendara el resultado. Cuando el vídeo terminaba se procedía a seguir con la reproducción automática, ya que así se puede ver la recomendación que hace el algoritmo de YouTube, y así hasta un número determinado de vídeos (con un máximo de 100 por tema). Por lo tanto, tenemos, al menos, dos primeros vídeos por tema, uno buscando con la cuenta y, el otro, sin ella. El objetivo es ver si, en ausencia de metadatos de cuenta, el algoritmo se comporta de forma diferente.

Los vídeos fueron analizados mediante la extracción de su contenido a través de los subtítulos (por lo tanto, es un análisis sobre el guion textual, no de las imágenes). Como es un recurso que no está presente en todos los vídeos de la plataforma se decidió aplicar el análisis a un vídeo de cada 10 en la lista de reproducción (o el que estuviera más próximo de forma ordinal) siempre que se contase con un número suficiente de vídeos (como se verá, en el caso de los vídeos capturados sin cuenta en la plataforma se ha extraído el texto de todos los que lo contenían). Con ese texto se procedió a un análisis de polarización afectiva (sobre el debate en torno al concepto véase Iyengar et al., 2019) mediante una técnica que ya hemos usado en ocasiones anteriores con bastante éxito (véase Serrano-Contreras et al., 2020, para una explicación pormenorizada). Básicamente, el procedimiento consiste en una modificación de un análisis de sentimiento (la herramienta seleccionada ha sido Orange3, Demsar et al., 2013, basada en Python y que utiliza un diccionario multilingüe para más de 50 idiomas). La modificación consiste en calcular la media de sentimiento de un conjunto de datos dado y medir la distancia entre la unidad de análisis y el global de la muestra (siendo, por lo tanto, un análisis de la distancia afectiva). En este caso hemos variado el análisis para medir la polarización de los vídeos y comentarios a partir del sentimiento del primer vídeo por tema y cuenta. Así, la polarización de los demás vídeos y comentarios se ofrece como la distancia con respecto al primero, pero en número absoluto. De esta forma podemos saber si un vídeo en concreto tiene un tono que se aleja (por positivo o negativo) del primero. No nos fijamos en lo negativo o positivo de los vídeos, ya que son consideraciones coyunturales y que exigirían un análisis pormenorizado del contenido, sino solo su distancia con respecto al primero. Entendemos que la polarización afectiva puede darse tanto mediante cargas positivas como negativas (por ejemplo, jaleando o felicitando a un grupo terrorista). El mismo análisis se hizo con respecto a los comentarios, con la salvedad de que se analizaron todos los posibles, incluyendo vídeos sin subtítulos (aunque no todos los vídeos permiten comentarios). Los resultados numéricos se encuentran en el rango de 0 a 100, pero es habitual obtener números muy bajos (en el entorno del 0,1,2). Esto es así porque la inmensa mayoría del contenido no está cargado afectivamente (incluso después de pre-procesar el texto con las técnicas habituales, tal y como se ha hecho). La única consecuencia es que cambios menores en el valor indican variaciones sustanciales en la polarización afectiva; después de todo hablamos de millones de palabras analizadas (Tabla 2).

De cada vídeo, entonces, se extraen las siguientes variables: aquellas provistas por la plataforma (número de me gusta, visualizaciones, número de comentarios, etc.), su tema, su posición en la lista de reproducción automática (recomendación de la plataforma), la polarización del contenido (uno de cada diez en los extraídos con cuenta) del vídeo y la polarización de los comentarios sobre el vídeo (en aquellos que los tienen habilitados). Las Tablas 1 y 2 describen el número de unidades de análisis (750 vídeos y, contando con los comentarios, cerca de tres millones de palabras).

	<b>Cuenta</b>	<b>Sin Cuenta</b>	<b>Total</b>
Comida vegetariana	101	6	107
COVID	100	5	105
Feminismo	101	11	112
Guerra en Ucrania	103	7	110
Podemos	89	4	93
Política nacional	108	0	108
VOX	109	6	115
<b>Total</b>	<b>711</b>	<b>39</b>	<b>750</b>

Nota. Posteriormente se analizaron de forma comparada los resultados del análisis de polarización sobre los vídeos y los comentarios. Como las hipótesis indican, se espera que cada vídeo, y los comentarios al mismo, en la lista de reproducción sea un poco más polarizante que el anterior.

### 3. Análisis y resultados

Un elemento que es necesario mencionar es que YouTube cambia sus parámetros y formas con el tiempo. Por ello, en esta investigación no pudimos trabajar tanto con los «me gusta» (likes), como con los «no me gusta», ya que ahora solo se ofrecen los datos de positividad. No obstante, no creemos que afecte a la investigación.

Siguiendo en la misma línea, YouTube no provee el mismo contenido a un usuario que a otro, algo que a tenor de lo ya reseñado tanto por Pariser (2017), como por el consumo cotidiano, es evidente. Sin embargo, ocurre otro hecho a tener en consideración. Entrar con o sin cuenta produce resultados distintos (Tabla 1). Pese a la cotidianidad de lo expuesto, existe otro elemento interesante que se ha alcanzado en este trabajo, cuando se efectúa una reproducción automática mediante su botón de continuar, si el proceso se realiza con cuenta, el modelo sigue ofreciendo vídeos, en cambio, al hacerlo sin ella, este termina entrando en un bucle en el que dos vídeos tienden a repetirse una y otra vez. De ahí la necesidad de poner un límite en el vídeo número 100, pero que es innecesario en el caso de que se acceda sin cuenta. Hecho, por otro lado, que también afecta a la duración de los vídeos. Mientras que sin cuenta la duración de estos no parecía un factor a tener en cuenta, con ella, y habiendo fijado un consumo constante, el modelo tendía a ofrecer vídeos cada vez más largos.

	Cuenta		Sin Cuenta		Total
	Subtítulos	Comentarios	Subtítulos	Comentarios	
Comida vegetariana	55.394	323.179	17.634	31.947	428.154
COVID	57.378	200.049	19.513	24.526	301.466
Feminismo	35.514	361.534	31.337	55.080	483.465
Guerra en Ucrania	14.989	221.559	2.324	16.327	255.199
Podemos	55.748	251.270	14.005	12.263	333.286
Política nacional	58.693	405.947	0*	0*	464.640
VOX	43.011	445.720	15.598	14.542	518.871
<b>Total</b>	<b>320.727</b>	<b>2.209.258</b>	<b>100.411</b>	<b>154.685</b>	<b>278.5081</b>

Nota. En el caso de la política nacional no hubo resultados coherentes, seguramente por la falta de metadatos sobre el país de origen de la búsqueda (se usó la red Tor para evitarlos).

Otro dato que considerar, y que ya está apuntando en la evolución de Davidson et al. (2010) a Covington et al. (2016), es que el ranking de resultados de YouTube está variando con el tiempo. Esto parece comprobarse en el tipo de contenido que ofrece la reproducción automática. Por ejemplo, si hablamos del tema de interés o de un canal con gran impacto. Esto se aprecia tanto en el tema del feminismo, como en el del vegetarianismo. Si bien la primera parece estar influenciado por haberse seleccionado en primera instancia un vídeo del canal de charlas TED, en la segunda, aun pudiendo generar posturas enfrentadas como el vegetarianismo, el tipo de vídeo, en este caso basado simplemente en recetas de cocina, ha hecho que el algoritmo no tendiese a otras vías. Como sí ocurrió con el resto de temáticas, que se bifurcaron y se desviaron a otros ámbitos. Señalar, por ejemplo, el caso del conflicto en Ucrania que gran parte de su muestra final son vídeos de música relajante. Así, hay factores que el algoritmo pretende premiar a fin de filtrar un determinado contenido que ofrecer al usuario. Postura que parecen apuestas claras de la compañía (véanse también Goodrow (2021) o Mohan (2022) para una explicación pormenorizada acerca de las actuaciones acometidas por la plataforma en aras de confeccionar unos contenidos menos nocivos tanto para la información como para el consumo de usuarios de todos los rangos de edad). De ahí que la Tabla 2 muestre datos que, en ocasiones, pueden resultar paradójicos, como que hay más palabras analizadas procedentes de subtítulos que de comentarios.

	Cuenta		Sin Cuenta	
	Subtítulos	Comentarios	Subtítulos	Comentarios
Comida vegetariana	0,23	0,81	0,11	0,14
COVID	1,84	2,13	0,18	1,00
Feminismo	0,67	0,96	0,45	0,48
Guerra en Ucrania	0,74	0,99	0,03	0,49
Podemos	0,21	1,29	0,23	0,11
Política nacional	0,56	0,30	-	-
VOX	0,96	0,41	0,21	2,16
$\bar{x}$	<b>0,76</b>	<b>0,91</b>	<b>0,22</b>	<b>0,67</b>

Si se observan los datos de polarización agregada media (Tabla 3) es fácil observar las diferencias en los valores según los temas, las cuentas y si proceden de los propios vídeos o de los comentarios. En el primer caso, hay claramente temas donde la polarización es más alta, sobre todo los referentes a la «COVID» ( $\bar{X}=1,35$ ), seguido de «VOX» (0,68), «política nacional» (0,56) y «feminismo» (0,55). Aunque los resultados no resultan sorprendentes, se observan importantes diferencias entre los analizados con cuenta y sin cuenta (sobre todo en el caso de «VOX» y de la «COVID»). Exceptuando en el caso de la «COVID» (2,07), los comentarios de los usuarios no muestran un patrón similar, siendo la «guerra de ucrania» (0,95), «podemos» (0,95) y «feminismo» (0,88), además del ya mencionado de la «COVID», los temas donde se ha observado una mayor polarización. Aunque también hay diferencias notables con respecto al origen del vídeo (con o sin cuenta), los datos parecen apuntar a la existencia de una posible cámara de eco en algunos casos o, al menos, de cierto grado de acuerdo entre los usuarios con cuenta que comentan en los vídeos que hemos catalogado como «VOX» y «política nacional», mientras que los que comentan en los vídeos capturados con las búsquedas «COVID», «podemos» o «guerra de ucrania», muestran signos de comentarios muy polarizantes.

En términos generales, los vídeos analizados con una cuenta de Google tienen una polarización media de 0,76 con una dispersión bastante alta (+- 0,70), mientras que los que se han analizado sin cuenta es mucho menor, de 0,22 (+-0,25). La distancia es menor si tenemos en cuenta la polarización en los comentarios de los usuarios, 0,91 y 0,67 respectivamente; en ambos casos sustancialmente mayor que la obtenida de los propios vídeos.

Como se puede apreciar en la Tabla 4, la matriz de correlaciones aporta datos muy interesantes. Por un lado, tenemos las correlaciones esperadas, tales como la relación entre likes, comentarios y visualizaciones; las correlaciones son muy fuertes porque, en realidad, están midiendo lo mismo: la popularidad de un vídeo determinado. Por otro lado, la variable que más nos interesa es Order, que nos indica el puesto en la lista de reproducción ofrecida por YouTube, siendo aquí donde hay hallazgos interesantes. Fundamentalmente podemos apreciar una correlación positiva entre la polarización y el orden del vídeo, tanto en los comentarios (Pol-Com) como en el propio vídeo (Pol-Sub) y, además, entre ellas. Es cierto que no nos encontramos ante una relación excesivamente fuerte, pero tampoco desdeñable, sobre todo la relación entre la polarización del vídeo con el orden (.316).

**Tabla 4. Matriz de correlaciones (Pearson)**

	ViewCount	Comments	Likes	Order	VisitCount	Pol-Com	Pol-Sub
ViewCount	1	.679**	.808**	-.069	-.113	-.064	.161
N	755	755	755	755	711	676	113
Comments	.679**	1	.786**	-.059	-.062	-.174	.045
N	755	755	755	755	711	676	113
Likes	.808**	.786**	1	-.027	-.053	-.146	-.076
N	755	755	755	755	711	676	113
Order	-.069	-.059	-.027	1	.132**	.124**	.316**
N	755	755	755	755	711	676	113
VisitCount	-.113**	-.062	-.053	.132**	1	-.082*	-.060
N	711	711	711	711	711	632	69
Pol-Com	-.064	-.174**	-.146**	.124**	-.082*	1	.265**
N	676	676	676	676	632	676	113
Pol-Sub	.161	.045	-.076	.316**	-.060	.265**	1
N	113	113	113	113	69	113	113

Nota. \*La correlación es significativa  $p < 0.01$  (2-colas). \*\*La correlación es significativa  $p < 0.05$  (2-colas).

Con respecto a la polarización de los vídeos, es interesante explorar dicha relación de forma más profunda. La relación entre ambas variables no es lineal (los resultados de intentar modelizarla mediante una regresión lineal han sido infructuosos,  $x^2 = -0,002$ ), por lo que es probable que haya un moderador. Para descubrir la variable que puede estar alterando esa relación se decidió analizar las medias, pero recodificando la variable Order en tres tramos (Order-Cat): de uno a tres vídeos, de cuatro a nueve y más de nueve. La segmentación no es, naturalmente, arbitraria. Hemos estimado que es posible, y hasta probable, que un usuario vea hasta tres vídeos seguidos propuestos por YouTube. Nos parece menos probable que vea entre cuatro y nueve vídeos, y bastante improbable que se vean más de nueve vídeos seguidos propuestos por el algoritmo. Así, consideramos que, en el primer caso, estaríamos ante una exposición ante el algoritmo baja, media en el segundo caso y alta en el tercero.

Tabla 5. Comparación de medias (solo se muestran las significativas)

	Order-Cat		
	(baja exp.)	(media exp.)	(alta exp.)
	Media	Media	Media
Com-Veg (Pol-Com)	0,11	0,26	0,86
Covid (Pol-Com)	0,41	1,65	-
Covid (Pol-Sub)	0,25	0,19	2,15
Feminismo (Pol-Com)	0,26	0,70	-
Ucrania (Pol-Com)	0,32	0,29	1,07
Pol-Nac (Pol-Sub)	0,20	-	0,63
Podemos (Pol-Com)	0,07	-	1,61
VOX (Pol-Com)	-	1,26	0,42
VOX (Pol-Sub)	0,12	0,26	1,07

Nota. Los resultados se basan en pruebas de dos partes asumiendo varianzas iguales. Las pruebas se ajustan para todas las comparaciones por pares utilizando la corrección de Bonferroni. Todos los niveles de significatividad son .05.

La Tabla 5 muestra los resultados de la comparación de medias, aunque solo en aquellos casos en que el test indica diferencias significativas. La primera impresión es que cada tema parece comportarse de una forma diferente: los vídeos (Pol-Sub) ofrecidos por la plataforma cuando se buscan los términos «COVID», «política nacional», y «VOX», efectivamente tienden a ser más polarizantes en cada uno de los tres tramos propuestos (exceptuando del primero al segundo de «COVID»). El caso de «VOX» es, quizá, el más claro, y muestra bien cómo la polarización media aumenta claramente en cada tramo de forma casi lineal. Con respecto a los comentarios, los temas donde se aprecian diferencias son «comida vegetariana», «COVID», «feminismo», «guerra de ucrania», «podemos» y «vox». Es decir, en todos los temas. Pero, al contrario que en el caso de los vídeos, sí se observan tendencias diferentes: mientras que en todos ellos la tendencia sigue siendo a un aumento de la polarización, en VOX se observa el comportamiento contrario.

Tanto en el caso de los vídeos como en el de los comentarios, esperábamos una relación similar entre los diferentes temas, pero las diferencias son tan sustanciales que difícilmente podemos validar las hipótesis de partida. Sí que parece que hay cierta relación entre la posición del vídeo —y sus comentarios— y un aumento de la polarización, pero las excepciones evitan que podamos afirmar que dicha relación es directa. Quizá algunos casos especialmente llamativos estén siendo modificados por la propia compañía, como con la COVID, donde se hizo un esfuerzo para que los ciudadanos recibiesen una información menos extrema. En cualquier caso, los resultados mostrados parecen indicar moderadores difíciles de medir en esa relación.

#### 4. Discusión y conclusiones

Las aproximaciones al debate del papel que juegan los algoritmos se están consolidando lentamente. Pese a ello, y si bien las variadas posturas comienzan a dejar claros trazos acerca de a qué nos podemos estar enfrentando, aún hoy hay otros aspectos que son difíciles de abordar. Este es el caso de los sistemas de recomendación (Yésilada & Lewandowsky, 2022). En este particular, trabajos como el presente, sirven para hacer un retrato acerca de qué ocurre en procesos como los de YouTube. Atestiguando, desde diferentes perspectivas, una realidad cambiante, pero de la que se pueden atesorar algunas reflexiones.

Parece evidenciarse, al menos mediante este modelo de muestreo, que las temáticas elegidas no tienden a extremarse. O, lo que es lo mismo, si se parte de un vídeo sobre comida vegetariana no se termina en uno sobre el veganismo o los movimientos «antiespecistas». Circunstancia que parece estar expuesta a factores más complejos que el simple tema de búsqueda, sin dejar de lado la capacidad que el modelo de entrenamiento de YouTube pueda tener para poder acceder a los datos e intereses del usuario y generar un patrón de consumo. Variable, a su vez, a tener en cuenta a la hora de ejecutar actuaciones de recogida de datos con esta serie de métodos de muestreo. Pese a ello, no podemos ofrecer una respuesta contundente. Como se ha mostrado al inicio, los teóricos establecen que los algoritmos de recomendación pueden estar provocando los fenómenos denominados burbujas de filtrado y cámaras de eco (Terren & Borge-Bravo, 2021).

Sin embargo, los resultados presentados aquí no son concluyentes: hay temas donde el algoritmo sí parece comportarse en esa dirección, pero hay otros temas donde esa relación, al menos con nuestra estrategia de investigación, no se percibe. No obstante, los resultados ofrecidos son llamativos y coherentes con investigaciones previas, donde comprobamos que el tema es un fuerte moderador de la polarización



de los contenidos y las reacciones de los usuarios (Serrano-Contreras et al., 2020). Esto abre otros interrogantes que son merecedores de una mayor atención investigadora: si los temas son un moderador en esa relación, ¿en qué temas podríamos afirmar que esas hipótesis se cumplen? Y, más importante, ¿por qué? Los datos ofrecidos aquí son insuficientes para responder a esas preguntas, más allá de la mera descripción de los escogidos, aunque aportan humildemente en la dirección de que la relación, de existir, no es directa ni lineal. Pero una selección basada en diferentes agrupaciones, por ejemplo, temas sobre política, extremismo, música, etc., además de implementar un seguimiento coherente de los clústeres de los denominados prosumidores. Y, con la adecuada coherencia interna, se podrían alcanzar conclusiones que caminen hacia la contestación definitiva a la pregunta que los teóricos llevan una década, o más, haciendo: ¿polarizan las redes sociales a nuestros ciudadanos?

## Notas

<sup>1</sup>Se ha de señalar que parte de las acciones ejecutadas para obtener las muestras se han fundamentado en la recapitulación metodológica, de un modo u otro, de trabajos empíricos previos, los cuales, en el caso de la reproducción automática, continuaban de forma sistemática el consumo. Esta aclaración se expone, principalmente, a que los autores consideramos que ese tipo de consumo mediático se aleja totalmente de la mayor parte de acciones que, en plataformas de vídeo -principalmente de contenido breve-, los usuarios llevan a cabo. Por lo que este tipo de acciones, basadas en un consumo constante y sin pausas o sin alteraciones en la periodicidad del mismo, afectan directamente a los resultados que el modelo algorítmico terminará por ofrecer.

<sup>2</sup>El presente trabajo se ha servido de varias incursiones para llevar a cabo la obtención de los datos de la muestra. Señalar, antes de las explicaciones pormenorizadas, que este tipo de análisis pueden realizarse también desde la API para acceder a sus servidores, pero precisa de una cuenta de acceso, lo cual fue descartado por considerarse un proceso poco orgánico para la recolección. En primer lugar, se procedió a realizar una búsqueda mediante la reproducción automática de vídeos y sin tener una cuenta vinculada, además de rechazar todo tipo de factores que pudieran alimentar el microtargeting. Por otro lado, y ante la escasa magnitud de datos que se obtenían de este, y buscando acudir a un proceso de cifrado aún mayor, se recurrió a las técnicas de anonimización por Onion mediante el uso de capas a través del modelo incógnito de Tor a través de Brave. Además, para añadir una capa más se hizo uso de la VPN que proporciona nuestra universidad. Habiendo realizado el mismo proceso, nos encontramos con la misma disyuntiva que la búsqueda previa, la falta de magnitud en la muestra, así como un cese de actividad ante la falta de interacción con la plataforma. Por tanto, en última instancia, como se ha señalado, se acometió el proceso a través de una cuenta de usuario sin actividad.

<sup>3</sup>Todas las búsquedas de las palabras clave se realizaron en minúscula y con los acentos correspondientes.

## Contribución de Autores

Idea, J.G.M., I.J.S.C.; Revisión de literatura (estado del arte), J.G.M., I.J.S.C.; Metodología, J.G.M., I.J.S.C.; Análisis de datos, J.G.M., I.J.S.C.; Resultados, J.G.M., I.J.S.C.; Discusión y conclusiones, J.G.M., I.J.S.C.; Redacción (borrador original), J.G.M., I.J.S.C.; Revisiones finales, J.G.M., I.J.S.C.; Diseño del Proyecto y patrocinios, J.G.M., I.J.S.C.

## Apoyos

Esta investigación es parte del proyecto de I+D+i PID2021-128272NB-I00 financiado por MCIN/AEI/10.13039/501100011033/FEDER "Una manera de hacer Europa".

## Referencias

- Alfano, M., Fard, A.E., Carter, J.A., Clutton, P., & Klein, C. (2021). Technologically scaffolded atypical cognition: The case of YouTube's recommender system. *Synthese*, 199(1-2), 835-858. <https://doi.org/10.1007/s11229-020-02724-x>
- Almagro, M., & Villanueva, N. (2021). Polarización y tecnologías de la Información: Radicales vs. extremistas. *Dilemata*, 34, 51-69. <https://bit.ly/38YwliH>
- Arceneaux, K., & Johnson, M. (2010). Does media fragmentation produce mass polarization? Selective exposure and a new era of minimal effects. In A. Campbell, & L. Martin (Eds.), *American Political Science Association 2010 Annual Meeting*. SSRN. <https://bit.ly/3M1e7jJ>
- Arias-Maldonado, M. (2016). La digitalización de la conversación pública: Redes sociales, afectividad política y democracia. *Revista de Estudios Políticos*, 173, 27-54. <https://doi.org/10.18042/cepc/rep.173.01>
- Bail, C.A. (2021). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press. <https://doi.org/10.1515/9780691216508>
- Banaji, S. (2013). Everyday racism and «My tram experience»: Emotion, civic performance and learning on YouTube. [El racismo cotidiano y «Mi experiencia en un tranvía»: emoción, comportamiento cívico y aprendizaje en YouTube]. *Comunicar*, 40, 69-78. <https://doi.org/10.3916/C40-2013-02-07>
- Barberá, P. (2020). Social media, echo chambers, and political polarization. In *Social media and democracy: The state of the field, prospects for reform* (pp. 34-55). Cambridge University Press. <https://doi.org/10.1017/9781108890960>
- Berners-Lee, T. (2000). *Tejiendo la red*. Siglo XXI de España. <https://bit.ly/3wZ1NMx>
- Berrocal-Gonzalo, S., Campos-Domínguez, E., & Redondo-García, M. (2014). Media prosumers in political communication: Politainment on YouTube. [Prosumidores mediáticos en la comunicación política: El «politainment» en YouTube]. *Comunicar*, 43, 65-72. <https://doi.org/10.3916/C43-2014-06>

- Bishop, S. (2018). Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. *Convergence*, 24, 69-84. <https://doi.org/10.1177/1354856517736978>
- Castells, M. (2001). *La era de la información: Economía, sociedad y cultura*. Alianza Editorial. <https://bit.ly/3LX118w>
- Chadwick, A. (2009). Web 2.0: New challenges for the study of e-democracy in an era of informational exuberance. *I/S: A Journal of Law and Policy for the Information Society*, 5(1), 9-41. <https://bit.ly/3MZopSH>
- Chen, A., Nyhan, B., Reifler, J., Robertson, R., & Wilson, C. (2021). *Exposure to alternative & extremist content on YouTube*. Anti-Defamation League. <https://bit.ly/3MZ19E9>
- Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. In S. Sen, & W. Geyer (Eds.), *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191-198). Association for Computing Machinery. <https://doi.org/10.1145/2959100.2959190>
- Davidson, J., Livingston, B., Sampath, D., Liebal, B., Liu, J., Nandy, P., Van-Vleet, T., Gargi, U., Gupta, S., He, Y., & Lambert, M. (2010). The YouTube video recommendation system. In X. Amatriain, M. Torrens, P. Resnick, & M. Zanker (Eds.), *Proceedings of the fourth ACM conference on Recommender Systems* (pp. 293-296). Association for Computing Machinery. <https://doi.org/10.1145/1864708.1864770>
- Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbonar, J., Zitnik, M., & Zupan, B. (2013). Orange: Data mining toolbox. *Python. The Journal of Machine Learning Research*, 14(1), 2349-2353. <https://bit.ly/3pMIPBR>
- Dimopoulos, G., Barlet-Ros, P., & Sanjuas-Cuxart, J. (2013). Analysis of YouTube user experience from passive measurements. In *Proceedings of the 9th International Conference on Network and Service Management (CNSM 2013)* (pp. 260-267). IEEE. <https://doi.org/10.1109/CNSM.2013.6727845>
- Goodrow, C. (2021). *On YouTube's recommendation system*. Blog YouTube. <https://bit.ly/3wVWAxhA>
- Habermas, J. (1981). *Historia y crítica de la opinión pública*. Gustavo Gili. <https://bit.ly/3O0JOv1>
- Hernández, E., Anduiza, E., & Rico, G. (2021). Affective polarization and the salience of elections. *Electoral Studies*, 69, 102203. <https://doi.org/10.1016/j.electstud.2020.102203>
- Howard, J.W. (2021). Extreme speech, democratic deliberation, and social media. In *The Oxford Handbook of Digital Ethics* (pp. 1-22). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198857815.013.10>
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S.J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129-146. <https://doi.org/10.1146/annurev-polisci-051117-073034>
- Latorre, M. (2022). *Historia de la Web, 1.0, 2.0, 3.0 y 4.0*. Blog Marino Latorre. <https://bit.ly/38un7QH>
- Lilleker, D.G., & Jackson, N. (2008). *Politicians and Web 2.0: The current bandwagon or changing the mindset?* [Conference]. Politics: Web 2.0 International Conference.
- Luengo, O., García-Marín, J., & Blasio, E. (2021). COVID-19 on YouTube: Debates and polarisation in the digital sphere. [COVID-19 en YouTube: Debates y polarización en la esfera digital]. *Comunicar*, 69, 9-19. <https://doi.org/10.3916/C69-2021-01>
- McLuhan, H.M. (1959). Myth and mass media. *Daedalus*, 88(2), 339-348. <https://bit.ly/3GtIs9v>
- Messina, J.P. (2022). *New directions in the ethics and politics of speech*. Routledge. <https://doi.org/10.4324/9781003240785>
- Mohan, N. (2022). *Inside responsibility: What's next on our misinfo efforts*. Blog YouTube. <https://bit.ly/38XAngS>
- Nielsen, R., & Fletcher, R. (2020). Democratic creative destruction? The Effect of a changing media landscape on democracy. In *Social media and democracy: The state of the field, prospects for reform* (pp. 139-162). Cambridge University Press. <https://doi.org/10.1017/9781108890960.008>
- O'Reilly, T., & Battelle, J. (2009). *Web squared: Web 2.0 five years on*. O'Reilly Media. <https://bit.ly/3wYLBuG>
- Pariser, E. (2017). *El filtro burbuja: Cómo la web decide lo que leemos y lo que pensamos*. Taurus. <https://bit.ly/3x0UyDX>
- Rasmussen, S.H.R., & Petersen, M. (2022). *From echo chambers to resonance chambers: How offline political events enter and are amplified in online networks*. PsyArXiv. <https://doi.org/10.31234/osf.io/vzu4q>
- Rekoff, M.G. (1985). On reverse engineering. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(2), 244-252. <https://doi.org/10.1109/TSMC.1985.6313354>
- Serrano-Contreras, I., García-Marín, J., & Luengo, O.G. (2020). Measuring online political dialogue: Does polarization trigger more deliberation? *Media and Communication*, 8, 63-72. <https://doi.org/10.17645/mac.v8i4.3149>
- Sunstein, C.R. (2007). *Republic.com 2.0*. Princeton University Press. <https://bit.ly/3a3YFG8>
- Terren, L., & Borge-Bravo, R. (2021). Echo chambers on social media: A systematic review of the literature. *Review of Communication Research*, 9, 99-118. <https://doi.org/10.12840/ISSN.2255-4165.028>
- Tufekci, Z. (2018). YouTube, the great radicalizer. *The New York Times*. <https://nyti.ms/38VtS2Y>
- Van-Bavel, J.J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences*, 25(11), 913-916. <https://doi.org/10.1016/j.tics.2021.07.013>
- Wigand, R., Wood, J., & Mande, D. (2010). *Taming the social network jungle: From Web 2.0 to social media*. [Conference]. AMCIS 2010 Proceedings. <https://bit.ly/3NJF3VVI>
- Yesilada, M., & Lewandowsky, S. (2022). Systematic review: YouTube recommendations and problematic content. *Internet Policy Review*, (1), 11-11. <https://doi.org/10.31234/osf.io/6pv5c>