

DETECCIÓN Y PREVENCIÓN DEL ODIO POR ISLAMOFOBIA EN ESPAÑA

WILLIAM GONZÁLEZ-BAQUERO

DIRECTOR
CARLOS ARCILA CALDERÓN

PLAN DE INVESTIGACIÓN
PROGRAMA DE DOCTORADO FORMACIÓN EN LA SOCIEDAD DEL CONOCIMIENTO
UNIVERSIDAD DE SALAMANCA

8 DE JUNIO DE 2023

INTRODUCCIÓN Y JUSTIFICACIÓN DEL TEMA OBJETO DE ESTUDIO

En los últimos años, las plataformas digitales, especialmente las redes sociales, se han convertido el lugar en donde se propagan y albergan discursos de odio más que nunca. Estos discursos difunden ideas y actitudes polarizadas, intolerantes y discriminatorias. Grupos y partidos políticos ultraconservadores y ultracatólicos han utilizado estas plataformas para difundir su ideología de manera cada vez más explícita, aprovechando la coyuntura de crisis económica, sanitaria, medioambiental y de valores. Este tipo de discursos intolerantes se dirigen a menudo hacia colectivos minoritarios con características específicas como el origen nacional y étnico, los rasgos fenotípicos, el género, la clase social, y la identidad cultural y religiosa. Por lo tanto, diferentes formas de discriminación se solapan en los discursos de odio, como el racismo, la xenofobia, la aporofobia y la discriminación por motivos culturales, simbólicos y religiosos.

En países mediterráneos como España, la islamofobia está en aumento como una de las principales formas de discriminación, especialmente debido a la llegada de inmigrantes, refugiados y solicitantes de asilo procedentes de Oriente Medio, que han aumentado en los últimos años debido a los conflictos en la región. Aunque la mayoría de los inmigrantes en España provienen de Latinoamérica, la discriminación y los conflictos culturales y religiosos son más comunes con los inmigrantes del norte de África y Oriente Medio, que hablan diferentes idiomas, suelen ser de clases sociales bajas y practican el islam, que a menudo está estigmatizado en los países occidentales como España. Teniendo en cuenta esto, el discurso de odio islamóforo tiene la característica predominante de ser interseccional, en donde confluyen odios de carácter racista, xenóforo, religioso, mauróforo y/o de género. Existen trabajos que brindan un acercamiento a la problemática, como por ejemplo el desarrollado por Fuentes-Lara y Arcila-Calderón (2022), en donde se estudian manifestaciones de odio islamóforas en la red social Twitter a partir de hashtags directamente relacionados con odio hacia esta comunidad, o el estudio realizado por Zamora et. al., (2021) en donde se analiza de manera cuantitativa el impacto de una campaña en contra de la islamofobia en Twitter, así como los trabajos que se vienen desarrollando por parte del Ministerio de Inclusión, Seguridad Social, y Migraciones a través del Observatorio Español del Racismo y la Xenofobia (OBERAXE), en los que se establecen indicadores para la medición de discursos de odio en línea. Sin embargo, aún no existen estudios centrados en analizar la conversación en línea en torno al islam o referida a la población musulmana, tratando de establecer un mapa que permita identificar los sentimientos predominantes, y las creencias que brindan una base para ejecutar este tipo de conducta intolerante en línea.

Tomando los planteamientos hechos por Amores et al. (2020), que proponen 4 encuadres a través de los cuáles suelen ser representados connotativamente los migrantes y los refugiados: el de normalización, el de victimización, el de carga y el de amenaza, que los representa como un peligro tanto para la seguridad, como para los valores culturales de las sociedades occidentales; y por otra parte, la Teoría de la Acción Razonada (Fishbein, 1967) en donde se plantea que la intención de una persona para llevar a cabo una conducta es una determinante directa para la ejecución de dicho comportamiento, por lo tanto, conociendo las actitudes sería posible predecir las conductas. Las actitudes, al igual que las normas subjetivas están conformadas por las creencias las cuales se forman mediante la asociación de dos o más aspectos determinados de la experiencia del individuo,

se establecerá un marco teórico desde el cual se realizará el análisis de diferentes contenidos publicados en redes sociales, potencialmente islamóforos, con el objetivo de crear una conceptualización de la intolerancia en redes sociales en contra de la población y la cultura musulmana.

Es así como, mediante este trabajo, se pretende hacer contribuciones en dos líneas principales: en primer lugar, un aporte a nivel teórico en el estudio y comprensión de los discursos acerca del Islam y la población musulmana en España, identificando creencias, actitudes, motivaciones, sentimientos, y planteando un modelo que permitirá identificar un patrón de amenaza-odio-agresión (Martínez et al., 2022) en el discurso islamóforo; y en segundo lugar un aporte a nivel metodológico mediante el uso de métodos computacionales, que permiten el análisis de grandes cantidades de datos, y de métodos experimentales que permitirán comparar y contrastar la información obtenida de los métodos computacionales.

Cabe resaltar que esta tesis está enmarcada dentro del proyecto europeo ECO (CERV-2022-EQUAL-101084984) en el cual se realizará la identificación y el análisis de los principales marcos conceptuales utilizados en el ecosistema en línea relacionados con el odio anti-musulmán (mapa de marcos conceptuales sobre la xenofobia y el racismo hacia la población musulmana). Los marcos se clasificarán en torno a narrativas específicas que construyen un informe de marcos que a su vez servirá como línea base general de nuestro estudio para identificar actitudes asociadas a cada encuadre desde un enfoque estructural, es decir, dimensión afectiva (sentimientos asociados), dimensión cognitiva (conocimientos y juicios asociados) y dimensión conductual (predisposición a la acción). de igual manera, esta tesis también está asociada al proyecto ISHADE concedido por la Fundación Pluralismo y Convivencia, proyecto en el cual se busca crear un detector automatizado de discurso islamóforo para redes sociales.

HIPÓTESIS DE TRABAJO Y PRINCIPALES OBJETIVOS A ALCANZAR

En este trabajo se busca explorar en las bases cognitivas que dan origen al discurso de odio islamóforo como conducta, partiendo de las creencias que dan lugar a las actitudes y a las normas subjetivas con respecto al tema estudiado, tomando como referencia la Teoría de la Acción Razonada. Teniendo en cuenta los trabajos previos realizados en torno a los discursos de odio, en donde los frames narrativos con los que son identificados refugiados y migrantes suelen ser negativos (Amores et al., 2020), y sumando a ello los estudios realizados por Fuentes-Lara y Arcila-Calderón (2022) y Zamora et. al., (2021) en donde se analizan diferentes manifestaciones islamóforas en Twitter, la literatura existente no brinda un claro panorama de la representación de la población musulmana en redes sociales, y del discurso de odio islamóforo con sus características particulares. Esto lleva a plantear las siguientes preguntas de investigación:

¿De qué manera es representada la comunidad musulmana en redes sociales?

¿Qué marcos connotativos son más frecuentes en la conversación en torno a la comunidad musulmana?

¿Cuáles son las creencias subyacentes a la actitud más predominantes en el discurso intolerante anti-musulmán?

Los marcos connotativos planteados desde teorías de la comunicación brindan una base para el análisis de los discursos intolerantes, sin embargo el estudio del odio en si mismo como una experiencia afectiva propia de los individuos pero con fuertes características grupales no ha sido tan estudiada en redes sociales desde una perspectiva psicológica. Tomando estos postulados y aplicándolos al caso específico de la islamofobia, y sumando a esto un modelo para profundizar en la amenaza como encuadre en donde se identifiquen los efectos diferenciales de las amenazas simbólicas y las amenazas realistas (Martínez et al., 2022), con el que se identifiquen las particularidades de este tipo de discurso intolerante, que en muchas ocasiones desemboca en agresiones físicas, se plantean las siguientes preguntas de investigación:

¿Con qué tipo de amenaza se asocia frecuentemente a la comunidad musulmana?

¿Qué mecanismos subyacentes están asociadas a las amenazas simbólicas y realistas en el discurso islamófobo?

¿Qué características intragrupal influyen en el discurso islamófobo?

De esta manera se plantea como objetivo principal de esta investigación conceptualizar el discurso de odio islamófobo, identificando creencias, actitudes, motivaciones y sentimientos, desde un modelo cognitivo, mediante el uso de metodologías computacionales y experimentales. Para la consecución de este objetivo, en primer lugar se realizará un análisis de mensajes de la red social Twitter (la elección de esta red social es por conveniencia, teniendo en cuenta las ventajas que ofrece frente a la descarga de datos) en donde se identificaran los diferentes marcos connotativos que subyacen a la conversación en torno al islam en Twitter, a su vez se identificaran las creencias subyacentes a la actitud en los mensajes que han sido clasificados como discurso intolerante de carácter islamófobo. Es así que se plantean los siguientes objetivos específicos:

- Identificar los marcos connotativos más predominantes en la conversación que gira en torno a la población musulmana en redes sociales.
- Identificar las creencias subyacentes a las actitudes en las cuales se basan el discurso islamófobo en redes sociales.

Tras alcanzar estos dos objetivos específicos se espera desarrollar diferentes modelos para la detección automática de diferentes formas de discurso de odio islamófobo, para de esta manera monitorizar los contenidos de odio que expresen lenguaje intolerante de tipo anti-musulman y elaborar un mapa de cada una de las diferentes formas de discurso de odio islamófobo en donde se presente más de una característica que pueda generar rechazo. Los marcos connotativos planteados desde teorías de la comunicación brindan una base para el análisis de los discursos intolerantes, sin embargo, el estudio del odio en si mismo como una experiencia afectiva propia de los individuos pero con fuertes características grupales no ha sido tan estudiada en redes sociales desde una perspectiva psicológica. Es así que a través de una encuesta experimental se buscará abordar los discursos intolerantes desde esta postura, partiendo de un modelo que permita indagar en aquellas manifestaciones de odio en donde la intolerancia desde se ámbito cognitivo (el prejuicio) y comportamental (discriminación y agresión) desembocan en discursos islamófobos. Este modelo busca profundizar en la percepción de la amenaza como predisponente para el odio

como fenómeno cognitivo-afectivo. Partiendo de este postulado, se plantea los siguientes objetivos específicos:

- Establecer un modelo de amenaza-odio-agresión que permita identificar las características de la amenaza como predisponente del odio hacia la comunidad musulmana.
- Identificar los mecanismos cognitivos y afectivos subyacentes a las amenazas ligadas al discurso islamóforo

METODOLOGÍA A UTILIZAR

Esta investigación se encuentra estructurada en tres fases, cada una respondiendo a bloques metodológicos y dando respuesta a los interrogantes previamente planteados. Durante la primera fase del desarrollo de esta investigación, se realizará una revisión bibliográfica siguiendo los parámetros establecidos por García-Peñalvo (2022) de la literatura sobre el discurso de odio islamóforo y anti-musulmán para consolidar un marco conceptual y teórico bajo el cual se hará, posteriormente, un análisis de contenido de los datos que se recolectarán posteriormente, apoyando el trabajo de campo con la literatura existente, identificando las variables de interés para una mejor comprensión de la problemática. El desarrollo de este marco conceptual es de suma importancia para el desarrollo de todo el trabajo, debido a que mediante este se establecerán las variables a través de las cuales se categorizarán las particularidades de la intolerancia en torno a lo identificado como musulmán.

Durante la segunda fase se realizará una selección de fuentes de contenido de odio e intolerancia, para analizar la estructura de la comunidad en la red social y los patrones de propagación que siguen los discursos de carácter islamóforo, los cuales se construirán mediante la recopilación de datos en tiempo real; las cantidades esperadas de datos por unidad de tiempo serán significativas, pero manejables. Además de esto, y para la gestión eficiente de los datos, se debe implementar un conjunto de técnicas de análisis para organizar el contenido recopilado. Con respecto a la captura de contenidos, la existencia de herramientas como Twitter Academic API (v2) y CrowdTangle para Facebook e Instagram, me permitirá acceder a una selección de contenidos de estas redes sociales más amplia y completa, lo que hace factible dotar al modelo de una amplia muestra de contenido de calidad que sería imposible de estudiar sólo con métodos manuales. Los métodos de análisis más comunes que se utilizarán en el marco se refieren a la estimación de fecha, hora y geolocalización, detección de idioma, detección de nivel de odio, toxicidad y análisis de sentimientos. Esto será posible gracias a técnicas de inteligencia artificial de vanguardia, incluido el aprendizaje profundo, como las redes neuronales recurrentes y el aprendizaje automático de última generación (Collobert et al., 2011; Thelwall et al., 2011; Karami et al., 2017, Grimmer y Stewart, 2017; Jacobi et al., 2016). Los modelos de aprendizaje automático se basarán en la tecnología Transformers con corpus pre-entrenados como BERT y GPT-3, que son las formas más avanzadas de representación del lenguaje natural que existen hoy en día. Los resultados, así como todas las observaciones realizadas a lo largo del desarrollo de la herramienta de detección y seguimiento, se utilizarán en combinación con un trabajo manual (análisis de contenido y etiquetado de ejemplos) y otro más cualitativo para comprender en profundidad las características de la propagación de diferentes formas de discurso islamóforo en línea. Los datos no estructurados

recogidos previamente se estudiarán cualitativamente mediante análisis del discurso, junto con los datos secundarios existentes sobre el tema, complementándolo con el análisis bibliométrico de literatura sobre este tema realizado previamente.

En la última fase de esta investigación se realizará un estudio experimental que buscará establecer un modelo de amenaza-odio-agresión que permita identificar las características de la amenaza como predisponente del odio hacia la comunidad musulmana, además de identificar los mecanismos cognitivos y afectivos subyacentes en las amenazas realistas y simbólicas ligadas al discurso islamóforo. Para la ejecución de esta parte experimental, se tomará una muestra probabilística polietápica, y se aplicará una encuesta experimental, previamente validada y distribuida a través de la herramienta Qualtrics. Este instrumento será una adaptación de la Escala de odio apasionado (PHS) desarrollada por Zeki y Romaya (2008) y usada para estudiar modelos de amenaza-odio por Martínez et. al., (2022).

MEDIOS Y RECURSOS MATERIALES DISPONIBLES

Esta investigación se desarrolla desde el Programa de Doctorado en Formación en la Sociedad del Conocimiento (García-Peñalvo, 2014; García-Peñalvo et al., 2019) de la Universidad de Salamanca, el cual se establece como un programa académico de carácter multidisciplinar. Desde su portal web <https://knowledgesociety.usal.es/>, se pueden apreciar los avances, en donde se incluyen publicaciones, estancias y asistencias a congresos, que se realizarán durante el desarrollo de este trabajo. A su vez, esta investigación se realizará desde el Observatorio de Contenidos Audiovisuales (www.ocausal.es) de la Universidad de Salamanca. Este grupo de investigación cuenta con amplia experiencia en el desarrollo de investigaciones, que pueden ser evidenciadas en la página web del grupo y en esta propuesta. El Observatorio de Contenidos Audiovisuales se encuentra ubicado en la Facultad de Ciencias Sociales de la Universidad de Salamanca, lo que le permite tener acceso al material de apoyo y el talento humano del Departamento de Sociología y Comunicación de la Universidad, además de contar con el equipamiento suficiente para abordar este proyecto (instalaciones, recursos informáticos y los materiales necesarios), así como los protocolos metodológicos utilizados para asegurar su adecuado desarrollo. Es importante resaltar que el director de esta tesis, tienen una amplia experiencia en el desarrollo de experimentos y de estudios de Big Data en comunicación.

Cabe resaltar que el estudiante actualmente está vinculado al proyecto ECO (CERV-2022-EQUAL-101084984) mediante convocatoria pública de la Universidad de Salamanca (código 2022/U126/01) en donde realizará la identificación y el análisis de los principales marcos conceptuales utilizados en el ecosistema en línea relacionados con el odio anti-musulmán (mapa de marcos conceptuales sobre la xenofobia y el racismo hacia la población musulmana).

Por último, los recursos específicos que se van a utilizar y que resultan gratuitos, licenciados por la universidad o ya adquiridos por el alumno son:

- Bases de Datos internacionales: Web of science (FECYT), Scopus.
- Base de datos de producción científica hispana: Dialnet.
- Plataformas de acceso a revistas científicas: Springer Link, Science Direct.

- Portal de libros electrónicos: CIELO
- Repositorio documental de la Universidad de Salamanca: Gredos.
- Buscador de contenido científico-académico: Google Scholar.
- Gestor bibliográfico: Mendeley.
- Paquetería de Microsoft Office: Microsoft Word, Microsoft Excel, Microsoft PowerPoint.
- Programa de análisis estadístico: IBM SPSS Statistics.
- Herramienta digital de elaboración y distribución de encuestas: Qualtrics.
- Lenguaje de programación: Python 2.7.
- Entorno de desarrollo integrado para Python: Pycharm.
- Interfaz de programación de aplicaciones API

PLANIFICACIÓN TEMPORAL AJUSTADA A TRES AÑOS

Durante los meses transcurridos del periodo académico 2022-2023, se planteó el problema, se diseñó la investigación y se dió inicio a la revisión bibliográfica. Cabe resaltar que durante este tiempo se ha procedido a publicar dos artículos (aceptados) 1 JCR y 1 Scopus, uno que es resultado de la investigación realizada en el marco del TFM y bajo en donde se realizó un acercamiento al estudio de las creencias subyacentes a los discursos de odio en redes sociales en contra de minorías, y otro directamente relacionado con el discurso islamófobo en redes sociales. También, se ha avanzado en el desarrollo de los instrumentos y técnicas que se usarán en la segunda fase de esta investigación y que se encuentran en sintonía con el proyecto Evaluando Campañas de Odio ECO (CERV-2022-EQUAL-101084984), como lo son la recolección de datos y la creación de libros de códigos para la clasificación manual de los mensajes con contenido intolerante frente a la población musulmana. Esta investigación está dividida en tres grandes fases, cada una brindando una base teórica para el desarrollo de la siguiente, encontrando así una estructura articulada y coherente.

Primer año

- 1.1. Análisis bibliométrico sobre el discurso intolerante de carácter islamófobo
- 1.2. Selección de fuentes de contenido de odio e intolerancia para analizar la estructura de la comunidad y los patrones de propagación
- 1.3. Diseño de técnicas de agrupación y monitoreo en tiempo real distribuidas a gran escala para recopilar y organizar contenido de odio

Segundo año

- 2.1. Desarrollo de modelos de difusión sensibles al tema y técnicas de aprendizaje de representación de red para identificar patrones de propagación de contenido de odio

REFERENCIAS

- Amores, J. J., Calderón, C. A., & Blanco-Herrero, D. (2020). Evolution of negative visual frames of immigrants and refugees in the main media of Southern Europe. *Profesional De La Informacion*. <https://doi.org/10.3145/epi.2020.nov.24>
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12, 2493-2537.
- Fishbein, M. (Ed.). (1967). *Readings in attitude theory and measurement*. Wiley.
- Fuentes-Lara, C., & Calderón, C. A. (2023). El discurso de odio islamófobo en las redes sociales. Un análisis de las actitudes ante la islamofobia en Twitter. *Revista Mediterránea de Comunicación: Mediterranean Journal of Communication*, 14(1), 225. <https://doi.org/10.14198/medcom.23044>
- García-Peñalvo, F. J. (2014). Formación en la sociedad del conocimiento, un programa de doctorado con una perspectiva interdisciplinar. *Education in the Knowledge Society*, 15(1), 4-9.
- García-Peñalvo, F. J. (2022). Developing robust state-of-the-art reports: Systematic Literature Reviews. *Education in the Knowledge Society*, 23, Article e28600. <https://doi.org/10.14201/eks.28600>
- García-Peñalvo, F. J., Rodríguez-Conde, M. J., Verdugo-Castro, S., & García-Holgado, A. (2019). Portal del Programa de Doctorado Formación en la Sociedad del Conocimiento. Reconocida con el I Premio de Buena Práctica en Calidad en la modalidad de Gestión. In A. Durán Ayago, N. Franco Pardo, & C. Frade Martínez (Eds.), *Buenas Prácticas en Calidad de la Universidad de Salamanca: Recopilación de las I Jornadas. REPOSITORIO DE BUENAS PRÁCTICAS (Recibidas desde marzo a septiembre de 2019)* (pp. 39-40). Salamanca, España: Ediciones Universidad de Salamanca.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297. <https://doi.org/10.1093/pan/mps028>
- Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital journalism*, 4(1), 89-106
- Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and Research: A Systematic Literature Review Through Text Mining. *IEEE Access*, 8, 67698-67717. <https://doi.org/10.1109/ACCESS.2020.2983656>
- Martínez, C. A., Van Prooijen, J., & Van Lange, P. A. M. (2022). A threat-based hate model: How symbolic and realistic threats underlie hate and aggression. *Journal of Experimental Social Psychology*, 103, 104393. <https://doi.org/10.1016/j.jesp.2022.104393>

- Medina, R., Clemente, P. G., & Martínez, J. M. A. (2021). Análisis del discurso de odio sobre la islamofobia en Twitter y su repercusión social en el caso de la campaña «Quítale las etiquetas al velo». *Anàlisi: Quaderns de Comunicació i Cultura*, 65, 1-19. <https://doi.org/10.5565/rev/analisi.3383>
- Observatorio Español del Racismo y la Xenofobia (2020). Informe sobre la intolerancia y la discriminación hacia los musulmanes en España. Obtenido de https://www.inclusion.gob.es/oberaxe/es/publicaciones/documentos/documento_0131.htm
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2011). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 62(2), 419.
- Zeki, S., & Romaya, J. P. (2008). Neural Correlates of Hate. *PLOS ONE*, 3(10), e3556. <https://doi.org/10.1371/journal.pone.0003556>