

MÁSTER PRESENCIAL EN CIBERSEGURIDAD – \_ªEd.

DESARROLLO DE UNA HERRAMIENTA ENFOCADA A DETECTAR REDES DE BOTS EN REDES SOCIALES

TFM elaborado por: David Fernández Peña

Tutor/a de TFM: Andrés Iglesias

- Madrid y fecha -

# ÍNDICE:

[ÍNDICE: 1](#_Toc1106106469)

[1. INTRODUCCIÓN. 2](#_Toc1176581712)

[2. OBJETIVOS. 3](#_Toc735751931)

[3. CONCLUSIONES. 3](#_Toc2029169389)

[4. BIBLIOGRAFÍA. 3](#_Toc1602013687)

[5. ANEXOS. 3](#_Toc1486052787)

# INTRODUCCIÓN.

El auge de las redes sociales ha transformado radicalmente la forma en que las personas se comunican e interactúan en línea. Plataformas como Twitter (ahora X), Instagram o Facebook han permitido el intercambio instantáneo de ideas, opiniones y noticias en tiempo real, creando un entorno dinámico de información (Ferrara et al., 2016). Sin embargo, este ecosistema también ha sido explotado por actores malintencionados que buscan manipular y distorsionar el flujo de información a través de la creación y uso de redes de bots.

En la última década, el crecimiento de las redes sociales ha revolucionado la forma en que las personas acceden a la información y se comunican entre sí, facilitando el flujo de ideas y noticias a nivel global. Entre estas plataformas, Twitter, se destaca por su capacidad para conectar a millones de usuarios en tiempo real, permitiendo la diseminación rápida de mensajes que pueden influir significativamente en la percepción pública y en la toma de decisiones (Ferrara et al., 2016). No obstante, este entorno dinámico y de rápida evolución también presenta un espacio fértil para la proliferación de **bots sociales.** (Ferrara, Varol, Davis, Menczer, & Flammini, 2016; Yang et al., 2023).

Los bots sociales son cuentas automatizadas que pueden imitar el comportamiento humano en las redes sociales, publicando contenido, interactuando con otros usuarios y, en muchos casos, promoviendo agendas específicas o difundiendo desinformación (Aïmeur, Amri, & Brassard, 2023; Yang et al., 2023).

Investigaciones previas han demostrado que los bots pueden influir en el discurso político, aumentar artificialmente la popularidad de ciertos temas y distorsionar la percepción pública (Ferrara et al., 2016). Además, estudios han revelado que, durante eventos críticos como elecciones o crisis internacionales, los bots han jugado un papel significativo en la creación y propagación de narrativas falsas (Shen et al., 2023).

Los bots sociales pueden dividirse en dos categorías principales: bots benignos y bots maliciosos. Los primeros, como los bots de servicio al cliente y los que ofrecen alertas de noticias, pueden mejorar la experiencia del usuario (Varol et al., 2017).

Un ejemplo notable de bots sociales benignos son los bots de alerta de desastres naturales, diseñados para proporcionar información en tiempo real sobre emergencias y eventos críticos, mejorando la experiencia del usuario y contribuyendo a la seguridad pública. Un caso particular es el bot **SF QuakeBot**, que publica automáticamente actualizaciones sobre terremotos en el área de la Bahía de San Francisco. Este bot monitorea la información de terremotos en tiempo real y genera publicaciones en Twitter para notificar a los usuarios cuando ocurre un sismo significativo (Ferrara et al., 2016). Otro ejemplo son los bots de atención al cliente, que permiten a las empresas responder preguntas comunes de manera inmediata, optimizando así la eficiencia del servicio y mejorando la satisfacción del cliente (Yang et al., 2023).

Estos bots benignos cumplen una función positiva al agilizar el flujo de información y ofrecer respuestas automatizadas a consultas de usuarios, desempeñando un papel importante en la mejora de la experiencia en redes sociales y en la reducción de los tiempos de respuesta en situaciones de alta demanda informativa (Aïmeur, Amri, & Brassard, 2023).

En contraste, los bots maliciosos representan una amenaza significativa, ya que pueden difundir desinformación, influir en discusiones políticas, manipular los mercados financieros y generar cámaras de eco que refuerzan sesgos cognitivos y polarización social (Ferrara et al., 2016). Durante eventos críticos, como elecciones o crisis políticas, se ha observado un aumento en la actividad de estos bots, promoviendo narrativas específicas para influir en la opinión pública, un fenómeno documentado en el contexto de la guerra entre Rusia y Ucrania, donde los bots fueron utilizados para amplificar mensajes a favor de ambos lados del conflicto (Shen et al., 2023).  
Uno de los casos más documentados es el uso de bots durante las elecciones presidenciales en Estados Unidos en 2016 y 2020, donde se desplegaron bots en Twitter para difundir noticias falsas y aumentar artificialmente la visibilidad de ciertos candidatos y temas controvertidos. Estos bots fueron diseñados para imitar cuentas reales y publicaban contenido polarizante, creando cámaras de eco que amplificaban la percepción de apoyo o rechazo hacia ciertos candidatos y temas políticos (Ferrara et al., 2016; Varol et al., 2017). Esta intervención artificial en el debate electoral representó una amenaza directa a la integridad de los procesos democráticos, al manipular la percepción pública y crear divisiones entre votantes.

Otro ejemplo destacado se dio durante la guerra entre Rusia y Ucrania en 2022, donde bots de ambos lados del conflicto difundieron información y desinformación de manera coordinada. Los bots prorrusos y proucranianos aumentaron la frecuencia de publicaciones que apoyaban las respectivas posiciones de cada bando, con el objetivo de influir en la percepción global del conflicto. Este uso estratégico de bots no solo sirvió para manipular la narrativa pública sobre el conflicto, sino que también contribuyó a la polarización de la audiencia en redes sociales, exacerbando el conflicto en la esfera digital y manipulando la opinión pública a través de desinformación y propaganda (Shen et al., 2023).

La detección de bots en redes sociales ha sido objeto de numerosos estudios que buscan desarrollar herramientas automatizadas capaces de diferenciar entre cuentas humanas y automatizadas. Entre los sistemas de detección más conocidos se encuentra **Botometer**, una herramienta basada en técnicas de aprendizaje supervisado que analiza una combinación de características de la cuenta, como la frecuencia de publicaciones, el contenido de los tweets y las interacciones en la red (Yang et al., 2023). A pesar de su popularidad, Botometer presenta limitaciones, especialmente al analizar cuentas con datos incompletos o cuando se aplica fuera del entorno de Twitter, lo que subraya la necesidad de métodos que se adapten a múltiples plataformas y contextos (Ferrara et al., 2016; Ng & Carley, 2023).

Otro enfoque reciente es el uso de modelos de **mezcla de expertos**, como el desarrollado en el sistema **BotBuster**, que aborda la limitación de datos incompletos dividiendo el análisis en varios "expertos" que procesan distintas características de las cuentas, como el nombre de usuario y el contenido de las publicaciones. Este método ha demostrado ser efectivo, alcanzando tasas de precisión superiores a otras herramientas tradicionales, y muestra la importancia de emplear técnicas flexibles y adaptativas para la detección de bots (Ng & Carley, 2023).

Además, algunos investigadores han explorado el análisis de **secuencias de clics y patrones de comportamiento**, evaluando la probabilidad de que un usuario sea un bot en función de su actividad temporal y su interacción con otros usuarios (Shi et al., 2019). Este enfoque se basa en la premisa de que los bots, aunque intentan imitar el comportamiento humano, siguen patrones más predecibles y menos contextuales que los usuarios reales, lo cual puede explotarse para su detección.

Pese a estos avances, el desarrollo de herramientas de detección de bots se enfrenta a desafíos constantes. La capacidad de los bots para evolucionar y adaptarse a las nuevas medidas de detección hace que las herramientas existentes queden obsoletas rápidamente (Yang et al., 2019). Asimismo, la creciente sofisticación de los bots que emplean procesamiento de lenguaje natural para generar contenido y responder de manera coherente a interacciones humanas plantea problemas adicionales, ya que imitan cada vez mejor las respuestas humanas, dificultando la diferenciación entre ambos (Aïmeur, Amri, & Brassard, 2023).

Twitter, debido a su estructura de red abierta y su influencia en la opinión pública, ha sido un terreno fértil para la proliferación de bots sociales. Estos bots no solo afectan la calidad de las interacciones en línea, sino que también pueden ser utilizados para manipular procesos democráticos, fomentar la visibilidad de ciertas cuentas o productos, difundir noticias falsas y amplificar la desinformación (Varol et al., 2017). Es por eso por lo que Twitter será la red social elegida para el desarrollo de la herramienta de detección que abarca este documento.

# OBJETIVO.

Este trabajo tiene como objetivo desarrollar una herramienta eficaz para la detección de redes de bots en Twitter. La importancia de esta herramienta radica en su capacidad para identificar patrones de comportamiento automatizados que se esconden tras la fachada de cuentas aparentemente legítimas. Utilizando una combinación de técnicas de Machine Learning, esta herramienta busca mejorar la precisión de la detección de bots (Shi et al., 2019).

Twitter es una plataforma ideal para el análisis de redes de bots debido a su gran volumen de datos, su accesibilidad a través de su API y una gran cantidad de datasets nacidos de trabajos similares en la plataforma. Sin embargo, también presenta retos significativos debido a la variabilidad de los datos disponibles y la evolución constante de los patrones de comportamiento de los bots (Aïmeur, Amri, & Brassard, 2023; Ng & Carley, 2023).

La presencia de bots en las redes sociales no solo afecta a la calidad de las interacciones en línea, sino que también representa una amenaza para la seguridad de la información y la estabilidad de las democracias. Los bots pueden manipular la percepción pública, influir en el mercado y erosionar la confianza en las instituciones y en las plataformas de comunicación (Ferrara et al., 2020; Varol et al., 2017). Esto subraya la necesidad urgente de desarrollar herramientas de detección robustas y adaptables que puedan identificar y mitigar la influencia de estos agentes maliciosos en tiempo real. En este contexto, el presente trabajo propone el desarrollo de una herramienta innovadora que, mediante el análisis de patrones de comportamiento y técnicas avanzadas de aprendizaje automático, permitirá identificar redes de bots en Twitter y mejorar así la seguridad y la transparencia de las interacciones en línea.

En resumen, este trabajo propone el desarrollo de una herramienta que, basada en algoritmos de Machine Learning, permita la identificación de redes de bots en Twitter, contribuyendo así a la mejora de la seguridad y la transparencia en las interacciones en línea.

# CONCLUSIONES.

# BIBLIOGRAFÍA.

* Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. Social Network Analysis and Mining, 13(30), 1-36. <https://doi.org/10.1007/s13278-023-01028-5>
* Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104. <https://doi.org/10.1145/2818717>
* Ng, L. H. X., & Carley, K. M. (2023). BotBuster: Multi-Platform Bot Detection Using A Mixture of Experts. Proceedings of the Seventeenth International AAAI Conference on Web and Social Media (ICWSM).
* Shi, P., Zhang, Z., & Choo, K. K. R. (2019). Detecting malicious social bots based on clickstream sequences. IEEE Access, 7, 28855-28869. <https://doi.org/10.1109/ACCESS.2019.2901864>
* Shen, F., Zhang, E., Ren, W., He, Y., Jia, Q., & Zhang, H. (2023). Examining the differences between human and bot social media accounts: A case study of the Russia-Ukraine War. First Monday, 28(2). <https://doi.org/10.5210/fm.v28i2.12777>
* Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. International AAAI Conference on Web and Social Media.
* Yang, K.-C., Varol, O., Nwala, A. C., Sayyadiharikandeh, M., Ferrara, E., Flammini, A., & Menczer, F. (2024). Social bots: Detection and challenges. In T. Yasseri (Ed.), Handbook of Computational Social Science. Edward Elgar Publishing Ltd.

# ANEXOS.