

# Verificación automática de hechos

## Pregunta de investigación

¿Cómo se pueden mejorar los sistemas de verificación automática de hechos para abordar eficazmente las limitaciones identificadas en el manejo de afirmaciones realistas?

## Definición del problema

El problema no resuelto en la verificación automática de hechos es la eficacia limitada de los sistemas actuales cuando se enfrentan a afirmaciones realistas. Estas afirmaciones pueden no ser lo suficientemente específicas o la evidencia disponible podría ser inconclusa, lo que resulta en una verificación inadecuada o errónea.

## Contexto

La era digital ha traído consigo un aumento en la velocidad y el alcance de la distribución de información. Aunque esto tiene numerosos beneficios, también facilita la rápida difusión de información errónea o intencionalmente falsa. Las redes sociales, en particular, ayudan a acelerar la propagación de esta desinformación, afectando a ámbitos que van desde la salud, como se pudo observar en la pandemia del COVID-19, a la política.

Para combatir la desinformación, la verificación de hechos se ha convertido en un elemento crucial. Sin embargo, el proceso de verificación manual es lento y no es posible igualar al rápido ritmo de propagación de las noticias falsas. Por ello, es necesario desarrollar sistemas inteligentes para asistir en este proceso.

## Estado del arte

Los sistemas automáticos basados en técnicas del procesamiento del lenguaje natural (NLP) han surgido como herramientas prometedoras para asistir en la verificación de hechos. Estos sistemas generalmente comparan afirmaciones textuales con documentos de gran reputación que sirven de evidencia para evaluar la veracidad de la afirmación [1], mostrando un rendimiento aceptable en estos casos. Sin embargo, estudios recientes [2,3,4] han identificado limitaciones en la eficacia de estos sistemas cuando se aplican a conjuntos de datos que contienen afirmaciones más realistas, ya que estas son a menudo vagas o se basan en evidencias inconclusas y pendientes por corroborar.

## Objetivo

El objetivo de esta investigación es comprender mejor los desafíos al aplicar la verificación automática de hechos en afirmaciones realistas y encontrar formas de abordar las deficiencias encontradas en la literatura actual.

# Alcance

Este estudio implementará modelos de aprendizaje profundo de última generación para la verificación automática de hechos, explorará y analizará las limitaciones de los enfoques existentes cuando se aplican a afirmaciones realistas, y diseñará, implementará y evaluará enfoques novedosos para la verificación automática de hechos.

## Metodología

Para esta investigación primero se llevará a cabo una revisión exhaustiva de la literatura existente, enfocándose en trabajos relacionados con la verificación automática de hechos, modelos de aprendizaje profundo aplicables y evaluación de sistemas de procesamiento del lenguaje natural. Además, se estudiarán conjuntos de datos existentes, como FEVER [1], para identificar sus limitaciones y potencial para mejorar la verificación de afirmaciones realistas.

Posteriormente, se implementarán y evaluarán varios modelos de aprendizaje profundo de última generación para la verificación automática de hechos. Considerando tanto enfoques supervisados como no supervisados. A continuación, se definirán métricas de evaluación, como precisión, exactitud y F1, y se realizarán experimentos para comparar el rendimiento de diferentes modelos y enfoques. Los modelos utilizados serán de código abierto e implementados en bibliotecas populares como PyTorch para asegurar la reproducibilidad.

Una vez extraídos los resultados de los experimentos, serán analizados para entender cómo los diferentes modelos y enfoques abordan las limitaciones en la verificación de afirmaciones realistas, discutiendo los hallazgos en el contexto de la literatura existente, resaltando las contribuciones y limitaciones del estudio.

Finalmente, se resumirán los hallazgos principales y se plantearán recomendaciones para investigaciones futuras en el campo de la verificación automática de hechos.

## Referencias

- [1] Thorne et al. "FEVER: a Large-scale Dataset for Fact Extraction and VERification". NAACL. 2018.
- [2] Hanselowski et al. "A Richly Annotated Corpus for Different Tasks in Automated Fact-Checking". CoNLL. 2019.
- [3] Diggelmann et al. "CLIMATE-FEVER: A Dataset for Verification of Real-World Climate Claims". Tackling Climate Change with Machine Learning Workshop at NeurIPS. 2020.
- [4] Thorne et al. "Evidence-based Verification for Real World Information Needs". CoRR. 2021.