FactCheckEC-Sci: detección de afirmaciones científicas en un corpus ecuatoriano de fact-checking en español

FactCheckEC-Sci: Detection of Scientific Claims in an Ecuadorian Spanish Fact-Checking Corpus

Mariuxi del Carmen Toapanta-Bernabé[1,2], Miguel Ángel García-Cumbreras[1], L. Alfonso Ureña-López[1], Camila Johana Chusan-Zambrano[2], Anthony Daniel Pluas-Mero[2]

[1]Universidad de Jaén, Jaén, España

[2]Universidad de Guayaquil, Guayaquil, Ecuador

mctb0005@red.ujaen.es, mariuxi.toapantab@ug.edu.ec, magc@ujaen.es, laurena@ujaen.es, camila.chusanz@ug.edu.ec, anthony.pluasm@ug.edu.ec

Autor de correspondencia: mariuxi.toapantab@ug.edu.ec

Resumen: Este trabajo presenta FactCheckEC-Sci, un recurso para detectar afirmaciones científicas en un corpus ecuatoriano de fact-checking en español. A partir de 6299 registros, se filtraron 1804 publicaciones candidatas y, tras una revisión manual, se consolidaron 419 instancias de contenido científico. El esquema de anotación contempla tres categorías no excluyentes: afirmación científica verificable, referencia científica y contexto de investigación. Sobre este recurso se evaluaron modelos basados en Transformers y una línea base TF-IDF y Regresión Logística; entre los modelos evaluados, BETO obtuvo el mejor desempeño. Además, se exploró la recuperación automática de literatura académica como apoyo para la verificación. Los resultados muestran el potencial del corpus para fortalecer las herramientas de fact-checking en español.

Palabras clave: fact-checking, corpus anotado, afirmaciones científicas, desinformación.

Abstract: This paper presents FactCheckEC-Sci, a resource for detecting scientific claims in an Ecuadorian Spanish fact-checking corpus. From 6299 records, 1804 candidate posts were filtered and manually reviewed, yielding 419 instances of scientific content. The annotation scheme considers three non-exclusive categories: verifiable scientific claim, scientific reference, and research context. Transformer-based models and a TF-IDF and Logistic Regression baseline were evaluated; among them, BETO achieved the best performance. In addition, automatic scholarly evidence retrieval was explored as support for verification. The results show the potential of the corpus to strengthen fact-checking tools in Spanish.

Keywords: fact-checking, annotated corpus, scientific claims, misinformation.

1 Introducción

La desinformación digital representa uno de los desafíos más relevantes para el procesamiento del lenguaje natural en escenarios heterogéneos y de alta variación discursiva. En el ámbito hispanohablante, el problema adquiere una complejidad adicional cuando los contenidos circulan en redes sociales, titulares periodísticos, publicaciones breves o piezas de verificación, y se ven fuertemente influidos por el contexto sociopolítico local. Este trabajo se sitúa en la intersección entre la construcción de corpus especializados para *fact-checking*, la detección automática de contenido científico y la evaluación de aproximaciones clásicas y basadas en modelos de lenguaje para el español en tareas de clasificación multietiqueta. La principal novedad de la propuesta no reside únicamente en la clasificación automática, sino en la adaptación de un esquema de detección de discurso científico a un corpus real de *fact-checking* en español ecuatoriano, construido a partir de verificaciones efectivas y, además, conectado con una fase posterior de recuperación de evidencia académica. Esta combinación permite avanzar hacia herramientas de verificación más finas y mejor alineadas con escenarios reales de apoyo al *fact-checking* científico.

Aunque en los últimos años se han pro-

puesto recursos y enfoques para la detección de desinformación, la verificación de afirmaciones y el análisis de credibilidad en español (Gómez-Adorno et al., 2021; Bonet-Jover, Sepúlveda-Torres, y Martínez-Barco, 2023), los trabajos centrados en corpus especializados construidos a partir de verificaciones reales del contexto ecuatoriano siguen siendo limitados. En este sentido, nuestro trabajo amplía la línea iniciada en el contexto ecuatoriano sobre la detección de desinformación en publicaciones verificadas de X (Toapanta Bernabe, Garcia-Cumbreras, y Urena-Lopez, 2024).

En este trabajo se presenta FactCheckEC-Sci, un recurso orientado a la detección de afirmaciones científicas en un corpus ecuatoriano de *fact-checking* en español. El punto de partida fue un corpus base de 6299 registros procedentes de fuentes de verificación. A partir de este conjunto se realizó un proceso de filtrado inicial que permitió seleccionar 1804 publicaciones candidatas, seguido de una revisión manual mediante la cual se consolidaron 419 instancias de contenido científico. Sobre este subconjunto se definió un esquema de anotación multietiqueta compuesto por tres categorías no excluyentes: *afirmación científica verificable*, *referencia científica* y *contexto de investigación*. Posteriormente, se evaluaron tanto una línea base clásica basada en TF-IDF y Regresión Logística como modelos basados en Transformers para el español, con el objetivo de determinar su capacidad para identificar este tipo de contenido, obteniéndose el mejor rendimiento con BETO.

En contraste con trabajos previos centrados en la detección general de desinformación en publicaciones verificadas de X del contexto ecuatoriano, este artículo se enfoca en un problema más específico: la identificación del discurso científico en un corpus real de *fact-checking* en español. La contribución principal no reside únicamente en la comparación de modelos, sino también en la construcción de un recurso especializado derivado de verificaciones efectivas, en la adaptación de un esquema multietiqueta para distinguir entre afirmación científica verificable, referencia científica y contexto de investigación, y en la exploración de la recuperabilidad documental de este tipo de contenido. Asimismo, aunque este recurso puede integrarse en flujos más amplios de verificación automatizada, el foco del presente trabajo recae en la construcción del corpus, su anotación y la evaluación del componente científico, no en la demostración del sistema completo.

2 Trabajo relacionado

La detección automática de desinformación y la verificación de afirmaciones se han consolidado como líneas de investigación activas en el ámbito del procesamiento del lenguaje natural. En términos generales, esta área ha evolucionado desde enfoques centrados en la clasificación binaria de noticias falsas hacia escenarios más complejos que incluyen la verificación de afirmaciones, la recuperación de evidencia, la evaluación de la credibilidad y el análisis de la consistencia entre afirmaciones y fuentes documentales (Gómez-Adorno et al., 2021; Bonet-Jover, Sepúlveda-Torres, y Martínez-Barco, 2023; Aloy-Mayo y Rosso, 2026). En este contexto, uno de los principales retos sigue siendo la disponibilidad de recursos anotados de calidad, especialmente para lenguas distintas del inglés y para dominios específicos.

En español, los avances en la detección de desinformación han estado marcados por la construcción de corpus especializados y la adaptación de modelos neuronales preentrenados. Diversos trabajos han mostrado que los modelos basados en Transformers, y en particular las variantes monolingües para el español, ofrecen ventajas relevantes frente a las aproximaciones tradicionales cuando la tarea depende de matices semánticos, de la variación discursiva y de señales contextuales. En este sentido, BETO se ha consolidado como una de las principales referencias para tareas de PLN en español (Cañete et al., 2023), mientras que modelos como DeBERTaV3 (He, Gao, y Chen, 2023) y XLM-T (Barbieri, Espinosa Anke, y Camacho-Collados, 2022) permiten contrastar el efecto de distintas arquitecturas y estrategias de preentrenamiento.

A diferencia de SciTweets, que se orienta a la detección del discurso científico en publicaciones de redes sociales en un marco general, el presente trabajo se centra en un subconjunto derivado de un corpus de *fact-checking* ecuatoriano, en el que el contenido científico aparece inserto en piezas ya sometidas a verificación. Esta diferencia desplaza el problema desde la simple identificación de contenido relacionado con la ciencia hacia un

escenario más aplicado, en el que las afirmaciones detectadas pueden integrarse en flujos posteriores de validación y de recuperación de evidencia.

Esta línea ha sido reforzada por la organización de la tarea *Scientific Web Discourse Processing* en CheckThat! 2025, donde la subtarea 4a plantea explícitamente la detección de tres dimensiones: presencia de una afirmación científica, referencia a una publicación o estudio científico y mención de entidades científicas, como universidades o investigadores (Hafid et al., 2025; CheckThat! Lab, 2025). Esta formulación es especialmente próxima al esquema adoptado en este artículo, ya que nuestra propuesta distingue entre *afirmación científica verificable*, *referencia científica* y *contexto de investigación*. No obstante, a diferencia de los recursos compartidos en ese marco, nuestro trabajo se centra en un corpus de *fact-checking* en español del contexto ecuatoriano, construido a partir de verificaciones reales.

La literatura reciente también ha mostrado que la verificación de afirmaciones científicas requiere recursos y estrategias diferenciadas respecto de otros tipos de contenido. A diferencia de las afirmaciones políticas o de actualidad general, las afirmaciones científicas suelen depender de conocimientos expertos, de fuentes bibliográficas y de relaciones semánticas más finas entre enunciados, hipótesis y evidencia documental. Por ello, varios enfoques combinan la clasificación textual con mecanismos de recuperación de literatura académica, el emparejamiento semántico o la priorización de documentos. Además, el uso de representaciones semánticas densas, como las derivadas de Sentence-BERT, se ha consolidado como una estrategia útil para el filtrado semántico, la búsqueda aproximada y la recuperación basada en similitud (Reimers y Gurevych, 2019).

Otra brecha relevante se relaciona con las dimensiones geográficas y sociolingüísticas de los datos. En español, varios recursos y estudios previos sobre desinformación se han construido a partir de tareas compartidas, colecciones periodísticas o contextos no centrados en el ecosistema ecuatoriano de verificación, como ocurre en FakeDeS y en otros esfuerzos orientados a la anotación de confiabilidad y la detección de desinformación en español (Gómez-Adorno et al., 2021; Bonet-Jover, Sepúlveda-Torres, y Martínez-Barco, 2023). A su vez, recursos como SciTweets responden a un marco más general e internacional de discurso científico en redes sociales, sin partir de verificaciones efectivas del contexto ecuatoriano (Hafid et al., 2022).

En el caso del español ecuatoriano, esta limitación resulta especialmente relevante, ya que el ecosistema de verificación combina contenidos provenientes de redes sociales, portales digitales y piezas de *fact-checking* con referencias directas a actores, instituciones y debates del contexto nacional. En esta línea, el trabajo de Toapanta Bernabé et al. (Toapanta Bernabe, Garcia-Cumbreras, y Urena-Lopez, 2024) constituye un antecedente cercano al abordar la detección de desinformación en publicaciones verificadas de X provenientes de Ecuador Chequea y de Ecuador Verifica. No obstante, el presente trabajo avanza hacia un recurso más específico, centrado en la detección de afirmaciones científicas en contenido ya verificado y en su posible conexión con evidencia académica recuperable.

3 Corpus y proceso de anotación

3.1 Corpus de partida

El recurso presentado en este trabajo se derivó de un corpus más amplio del ecosistema ecuatoriano de *fact-checking*, desarrollado en el marco de una línea de investigación sobre la verificación automatizada de contenidos en español ecuatoriano. Esta línea se articula con el proyecto de investigación FCI-079-2023 y con desarrollos previos del sistema de verificación, entre los cuales se han generado distintos productos y recursos asociados. En este artículo, sin embargo, el objetivo no es describir exhaustivamente todo ese ecosistema, sino presentar el subconjunto, el esquema de anotación y los experimentos seleccionados específicamente para la detección de afirmaciones científicas en contenido verificado del contexto ecuatoriano.

Para las fases experimentales se utilizó una versión estructurada y enriquecida del corpus base, obtenida a partir de una versión cruda previa mediante procesos de transformación y normalización orientados a preservar la trazabilidad del documento original y a facilitar su explotación en tareas de anotación y modelado. En la versión enviada del manuscrito se mencionaban nombres internos de trabajo como `corpus_augmented.json` y `corpus_crudos.json`; en esta versión revisa-

da se explicitan como versiones internas del flujo de procesamiento del corpus, y no como denominaciones autónomas destinadas al lector final.

Los recursos asociados al corpus y la guía de etiquetado están disponibles para fines de consulta y reproducibilidad en un repositorio público.[1]

La estructura del corpus incluye, entre otros, los atributos `_uid`, `fuente`, `texto`, `titulo`, `fecha_noticia`, `etiqueta`, `raw`, `region` y `augmented`. En particular, el campo `texto` constituye la unidad principal de análisis para la tarea propuesta, mientras que `etiqueta` conserva el veredicto de veracidad del sistema original y `raw` permite recuperar el documento anidado con fines de trazabilidad y revisión contextual.

A diferencia de otros recursos construidos específicamente para la detección de discurso científico, el corpus de partida no incluía una variable objetivo asociada a las afirmaciones científicas. Por ello, fue necesario definir una nueva capa de anotación manual que permitiera identificar este fenómeno en publicaciones ya verificadas en el contexto ecuatoriano. Esta decisión metodológica resulta relevante porque transforma un corpus inicialmente orientado al *fact-checking* general en un recurso especializado para la detección multietiqueta de contenido científico.

3.2 Esquema de anotación

Para la construcción del subconjunto especializado se adoptó un esquema de anotación inspirado en el marco propuesto en SciTweets y adaptado al contexto de verificación en español. Este enfoque parte de una idea central: la relación de un texto con la ciencia no debe modelarse como una categoría binaria simple, sino mediante dimensiones diferenciadas que permitan distinguir entre afirmaciones verificables, referencias a estudios y menciones al entorno científico (Hafid et al., 2022). En esta línea, la formulación también resulta coherente con la orientación de la tarea *Scientific Web Discourse Processing* de CheckThat!, en la que se distinguen explícitamente las afirmaciones científicas, las referencias a publicaciones y las menciones de entidades científicas (Hafid et al., 2025; CheckThat! Lab, 2025).

El objetivo no fue clasificar el contenido científico únicamente por su tema, sino por su función discursiva en el proceso de verificación. Estas tres categorías se consideraron suficientes porque cubren funciones complementarias del discurso científico dentro del *fact-checking*: una afirmación contrastable, una referencia a la fuente y el contexto de investigación. Por ello, los casos anecdóticos, administrativos, políticos, vagos o no contrastables científicamente no se modelaron como una categoría positiva independiente. En consecuencia, se definieron tres categorías no excluyentes, de modo que una misma publicación puede recibir varias etiquetas simultáneamente:

- **Afirmación científica verificable:** se asigna cuando el texto contiene una proposición fáctica que puede contrastarse con evidencia científica, como artículos académicos, ensayos clínicos o estadísticas derivadas de la investigación. La verificabilidad se entiende aquí como la posibilidad teórica de validar la afirmación mediante documentos producidos por científicos o por instituciones de investigación (Hafid et al., 2022).
- **Referencia científica:** se asigna cuando la publicación proporciona un mecanismo de acceso, directo o indirecto, a la fuente del conocimiento científico. Esto incluye el DOI, la URL, el título exacto de una publicación o formas equivalentes de referencia a un estudio específico.
- **Contexto de investigación:** se asigna cuando el texto menciona explícitamente elementos del entorno académico o del proceso científico, tales como científicos, instituciones académicas, laboratorios, esfuerzos de investigación o hallazgos, aun cuando no formule una afirmación estrictamente verificable.

Este diseño multietiqueta permite representar con mayor precisión la complejidad del discurso científico en redes y en contenidos verificados. En particular, evita reducir todo texto que incluya terminología técnica a "científico" y distingue entre publicaciones que expresan un hallazgo verificable, aquellas que remiten a una fuente y aquellas que solo sitúan el contenido dentro de un marco de investigación.

3.3 Criterios operativos y resolución de ambigüedades

Dado que el discurso científico en redes sociales presenta formulaciones breves, elípticas y

[1] Véase: https://huggingface.co/datasets/mariuxi-toapanta/FactCheckEC-Sci

a menudo ambiguas, la anotación requirió criterios operativos adicionales para asegurar la consistencia en los casos límite. Para ello, se definieron reglas para resolver ambigüedades que complementaron las definiciones generales de las etiquetas.

En primer lugar, se aplicó una regla de *entidad vs. contexto*: la sola presencia de terminología científica o de entidades asociadas al ámbito sanitario o académico no fue considerada suficiente para clasificar una publicación como científica. Por ejemplo, las menciones a virus, hospitales o ministerios no implican, por sí mismas, una afirmación científica si el texto se refiere únicamente a decisiones administrativas, políticas o de gestión.

En segundo lugar, para la categoría *referencia científica* se contempló una regla de *referencia deíctica*. En consecuencia, se consideraron válidos ciertos casos en los que no aparecía una URL explícita en el texto plano, pero sí un señalamiento inequívoco a un contenido adjunto, como un estudio compartido mediante una tarjeta de previsualización, un *quote tweet* o una referencia demostrativa equivalente.

En tercer lugar, se utilizó una regla de *completitud del hallazgo* para distinguir entre *afirmación científica verificable* y *contexto de investigación*. Así, cuando una publicación mencionaba que "científicos descubrieron algo" o que "un estudio encontró resultados" sin explicitar el hallazgo concreto, se etiquetó únicamente como contexto y no como una afirmación verificable, debido a la vaguedad del contenido, que impedía su contraste eficaz con la evidencia científica.

Finalmente, se distinguió entre *autoridad política* y *autoridad científica*. Si una figura política emitía una afirmación verificable sobre ciencia, el contenido podía marcarse como afirmación científica, pero no como contexto de investigación salvo que se citara explícitamente un estudio, una institución científica o un actor del ecosistema académico.

Estas reglas permitieron reducir la arbitrariedad en la anotación y reforzaron la consistencia del recurso en los casos en que la lectura superficial del texto podía conducir a clasificaciones erróneas.

3.4 Filtrado preliminar y revisión manual

El proceso de anotación se inició sobre el conjunto `corpus_augmented`, compuesto por 6299 registros. Debido al volumen del corpus, no resultaba operativo revisar exhaustivamente todas las entradas desde el inicio. Por ello, se aplicó una fase de filtrado preliminar orientada a maximizar la sensibilidad y reducir el riesgo de excluir publicaciones potencialmente relevantes.

Este filtrado inicial combinó diccionarios de palabras clave con medidas de similitud semántica basadas en Sentence Transformers (Reimers y Gurevych, 2019), con el fin de discriminar entre contenido presumiblemente irrelevante y publicaciones con mayor probabilidad de contener discurso científico. Como resultado de esta etapa, se identificaron 1804 registros candidatos.

Posteriormente, los 1804 casos fueron sometidos a una revisión manual exhaustiva. En esta fase, cada publicación fue analizada según las tres categorías previamente definidas, determinando si el texto correspondía efectivamente a alguna manifestación del discurso científico. La naturaleza permisiva del filtro preliminar generó un número considerable de falsos positivos, lo cual era esperable, dado que la prioridad metodológica consistía en no perder cobertura en la etapa inicial.

Tras la depuración manual, se consolidó un conjunto final de 419 registros confirmados como afirmaciones científicas o publicaciones asociadas al discurso científico, mientras que 1385 registros fueron descartados por no encajar en ninguna de las categorías establecidas. Este proceso permitió obtener un subconjunto especializado, con mayor precisión semántica y mayor adecuación para el entrenamiento posterior de modelos de clasificación.

Con el fin de verificar que el filtro preliminar no hubiera excluido sistemáticamente contenido relevante, se realizó, además, una auditoría del conjunto de registros descartados. Para ello se calculó una muestra representativa mediante la fórmula de Cochran para poblaciones finitas (Cochran, 1977), obteniéndose una muestra aleatoria de 354 registros. Esta verificación adicional permitió reforzar la validez del procedimiento de selección y aportar mayor robustez metodológica al recurso resultante.

3.5 Control de calidad de la anotación

Dado que el corpus de partida no incluía etiquetas predefinidas relacionadas con el discurso científico, la construcción del recurso

requirió un proceso de revisión manual, guiado por definiciones operativas explícitas para cada categoría. En particular, se establecieron criterios diferenciados para distinguir entre *afirmación científica verificable*, *referencia científica* y *contexto de investigación*, evitando reducir el fenómeno a una clasificación binaria simplificada. Esta decisión permitió representar con mayor fidelidad la complejidad del discurso científico en publicaciones verificadas.

La revisión manual del subconjunto candidato se realizó siguiendo una guía de etiquetado común y definiciones operativas explícitas para cada categoría. Los casos ambiguos se resolvieron mediante la revisión contextual del contenido y la aplicación coherente de los criterios definidos.

Además, el proceso se apoyó en una guía de etiquetado elaborada para este recurso,[2] con criterios positivos y negativos, así como reglas de desambiguación, lo que permitió mantener la coherencia entre las afirmaciones, las referencias indirectas y las menciones al entorno científico.

En esta versión del artículo no se reporta una medida formal de acuerdo interanotador, ya que el énfasis metodológico se centra en la definición operativa de categorías, la revisión contextual de ambigüedades y la auditoría adicional del conjunto descartado como mecanismos de control de calidad del recurso. La Tabla 1 sintetiza las principales etapas cuantitativas del proceso de construcción del subconjunto especializado.

Fase	**Cantidad**
Corpus base	6299
Candidatos tras filtrado	1804
Instancias confirmadas	419
Descartados tras revisión	1385
Muestra auditada	354

Tabla 1: Resumen cuantitativo de la construcción de FactCheckEC-Sci.

3.6 Recurso resultante

La Tabla 2 presenta ejemplos representativos del esquema de anotación multietiqueta utilizado en FactCheckEC-Sci.

El resultado de este proceso es un corpus especializado para detectar afirmaciones científicas en publicaciones verificadas del contexto ecuatoriano. Desde la perspectiva de PLN, el recurso combina tres ventajas principales: se apoya en datos reales procedentes de un sistema de *fact-checking* ya existente, incorpora una anotación multietiqueta que permite modelar dimensiones complementarias del discurso científico y mantiene la trazabilidad al documento original y a los metadatos del sistema. Además, este subconjunto sirve de base para dos etapas posteriores del trabajo: la evaluación de modelos para la detección multietiqueta de afirmaciones científicas y la recuperación de evidencia académica cuando la publicación alude a un estudio o documento identificable con suficiente precisión. De forma complementaria, los registros descartados durante la depuración manual pueden interpretarse como material negativo o fuera de alcance útil para etapas previas de cribado, aunque no forman parte del recurso anotado principal presentado en este artículo.

Texto	**C1**	**C2**	**C3**
Los gatos tienen 7 vidas.	1	0	0
Estudio: los gatos viven más.	1	0	1
Estudio sobre gatos (enlace).	0	1	1
Científicos analizan COVID-19.	0	0	1

Tabla 2: Ejemplos del esquema de anotación multietiqueta.

4 Metodología

4.1 Diseño experimental

La metodología propuesta se estructuró en dos fases complementarias. La primera se orientó a la detección automática de afirmaciones científicas mediante la clasificación multietiqueta del subconjunto anotado manualmente. La segunda se centró en la recuperación de evidencia académica para aquellos casos en los que la publicación hacía referencia de forma suficientemente precisa a un estudio o documento fuente.

Este diseño responde a una lógica de pipeline propia del *fact-checking* automatizado: primero se verifica si una publicación contiene contenido científico relevante y, posteriormente, se intenta localizar literatura académica que respalde, contextualice o refute dicha afirmación. Aunque ambas fases están relacionadas, el foco principal de este artículo recae en la primera. La Figura 1 resume el flujo completo del proceso, desde el corpus base y el filtrado preliminar hasta la clasificación automática y la recuperación de evidencia académica.

[2] La guía de etiquetado utilizada en la fase de anotación está disponible en: https://huggingface.co/datasets/mariuxitoapanta/FactCheckEC-Sci

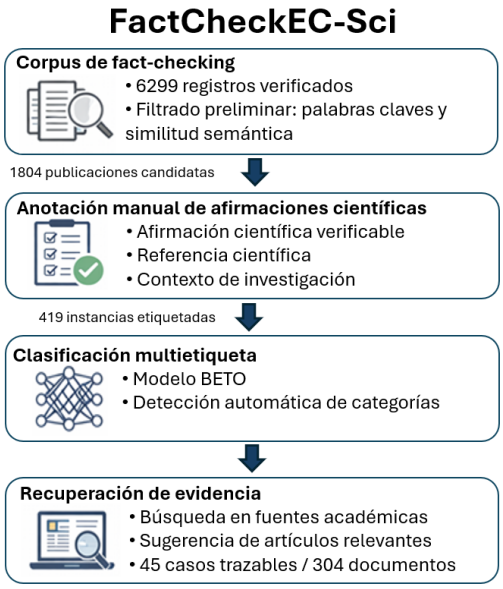

Figura 1: Flujo metodológico de FactCheckEC-Sci.

4.2 Preparación de datos

Antes del entrenamiento de los modelos, el conjunto de datos fue sometido a un proceso de limpieza y normalización textual. Este proceso incluyó la eliminación de saltos de línea, tabulaciones, espacios redundantes y otros caracteres no deseados que pudieran introducir ruido durante el aprendizaje. Asimismo, se preservó el contenido semántico principal de cada publicación, dado que la tarea depende de señales léxicas y contextuales específicas asociadas al discurso científico.

Una vez completada la limpieza, cada registro quedó representado por dos elementos principales: el texto de entrada y su vector binario de etiquetas. Este vector multietiqueta representa la pertenencia simultánea a las categorías definidas en la fase de anotación: *afirmación científica verificable*, *referencia científica* y *contexto de investigación*.

4.3 Partición del corpus y tratamiento del desbalance

Para la evaluación comparativa de los modelos, se aplicó una partición estratificada del corpus en conjuntos de entrenamiento (80 %) y de prueba (20 %). Esta estrategia permitió preservar, en la medida de lo posible, la distribución de las combinaciones de etiquetas menos frecuentes en ambos subconjuntos, reduciendo así el riesgo de sesgos en la evaluación.

Durante el análisis exploratorio se observó un marcado desbalance entre categorías. En particular, la categoría *referencia científica* no presentó instancias positivas suficientes para sostener un proceso de entrenamiento y evaluación fiable bajo el mismo esquema experimental que las demás. Por esta razón, la comparación de modelos se centró en las categorías *afirmación científica verificable* y *contexto de investigación*, manteniendo la segunda categoría como parte del esquema conceptual del recurso, pero no en la evaluación cuantitativa principal.

Aunque enfoques como la expansión iterativa del corpus con apoyo humano (*human-in-the-loop*), el sobremuestreo o la generación sintética de ejemplos podrían fortalecer las categorías minoritarias, en esta versión del trabajo se optó por no aplicarlos debido al tamaño reducido y a la fragilidad semántica de la categoría *referencia científica*, ya que podían amplificar el ruido, sesgar la distribución original o comprometer la fidelidad discursiva del recurso. En su lugar, se priorizó la evaluación de etiquetas con soporte empírico suficiente y, para mitigar los efectos del desbalance en las categorías finalmente evaluadas, se utilizó un enfoque de aprendizaje sensible al costo mediante el ajuste de los pesos de la función de pérdida (*loss weights*), otorgando mayor relevancia a las etiquetas minoritarias sin alterar artificialmente la distribución original del conjunto de datos.

Este escenario debe interpretarse, además, en relación con el propio proceso de construcción del recurso. El subconjunto final de 419 instancias no surge de una selección arbitraria, sino de un filtrado progresivo y de una revisión manual deliberadamente conservadora a partir de un corpus base mucho más amplio. En este sentido, la baja representación de *referencia científica* no solo constituye una limitación para la evaluación cuantitativa, sino también una característica empírica del dominio analizado, en el que las publicaciones verificadas aluden con frecuencia a la ciencia de forma narrativa o indirecta, sin aportar anclajes documentales explícitos. Por ello, la categoría se mantiene como parte del esquema multietiqueta del recurso, aunque su ampliación y fortalecimiento se plantean como una línea prioritaria de trabajo futuro.

4.4 Modelos evaluados

Se evaluaron cuatro configuraciones de clasificación con distintos niveles de complejidad y especialización lingüística: una línea base tradicional basada en TF-IDF y Regre-

sión Logística; DeBERTa-v3-base como arquitectura Transformer generalista; Twitter-XLM-RoBERTa-base como modelo multilingüe adaptado al dominio de redes sociales; y BETO como modelo monolingüe preentrenado para español. Esta selección permitió comparar no solo el rendimiento absoluto, sino también el efecto de la especialización lingüística en una tarea de clasificación en español ecuatoriano dentro del dominio de *fact-checking*. BETO, que obtuvo el mejor desempeño en la comparación experimental, se ajustó mediante *fine-tuning* para la tarea de clasificación multietiqueta utilizando la biblioteca Hugging Face Transformers.

4.5 Métricas de evaluación

La evaluación del componente de detección se realizó mediante la precisión (*precision*), la exhaustividad (*recall*) y el F1-score. Dado el desbalance entre clases, el indicador principal utilizado para la comparación global fue el Macro F1, por ofrecer una medida más equilibrada del rendimiento entre categorías y penalizar de manera más explícita los errores en las clases menos frecuentes. La Tabla 4 incorpora, además, las métricas por categoría para ofrecer una visión más completa del desempeño comparativo entre las configuraciones.

4.6 Recuperación de evidencia académica

Como extensión del recurso anotado, se implementó un procedimiento de recuperación de fuentes académicas para aquellas publicaciones cuya formulación permitía asociarlas a un estudio o documento fuente identificable. Para ello se adoptó un protocolo de correspondencia noticia-documento orientado a determinar, para cada publicación, si el contenido remitía a un único estudio y si los documentos recuperados mantenían una relación semántica suficiente con la afirmación expresada en el texto. Este enfoque permitió abordar la recuperación de evidencia no solo como una consulta bibliográfica, sino también como una tarea de emparejamiento entre una publicación verificada y un documento científico candidato.

En términos metodológicos, esta fase presupone una condición de elegibilidad: que la publicación aluda a un único estudio o a una fuente documental suficientemente delimitada, de modo que la recuperación pueda formularse como un problema de correspondencia entre noticia y documento, y no como una búsqueda abierta sobre un tema científico general.

La aplicación de este criterio permitió identificar 45 afirmaciones con suficiente especificidad documental para iniciar un proceso automático de recuperación de evidencia. Para este subconjunto se implementó un módulo de extracción en Python que consultó simultáneamente la API REST de Crossref y las E-utilities de PubMed (Crossref, 2026; National Center for Biotechnology Information, 2022). Para cada afirmación, se generó una consulta basada en palabras clave derivadas del contenido textual, que incluían nombres de instituciones, revistas, estudios o elementos distintivos del enunciado. Posteriormente, los resultados se contrastaron mediante metadatos como el título, el resumen, la fecha, las afiliaciones y la sede de publicación, a fin de reforzar la correspondencia entre la publicación del corpus y el documento científico recuperado. La Tabla 3 resume la

Elemento	**Cantidad**
Afirmaciones confirmadas	419
Casos trazables	45
Casos no trazables	374
Documentos candidatos	304

Tabla 3: Resumen de la recuperación de evidencia académica.

fase de recuperación de evidencia académica, distinguiendo entre el subconjunto trazable y el volumen final de documentos candidatos obtenidos. Esta fase refuerza el valor aplicado del recurso, ya que integra la detección de afirmaciones científicas en un flujo de verificación más amplio y aproxima escenarios reales de apoyo al *fact-checking* científico.

5 Resultados y discusión

5.1 Resultados de la detección de afirmaciones científicas

La Tabla 4 resume el desempeño de los modelos evaluados en las categorías de afirmación científica verificable (C1) y de contexto de investigación (C3). Además del Macro F1 como indicador global, se reportan el F1-score, la exhaustividad (Exh.) y la precisión (Prec.) por categoría. BETO obtuvo el mejor desempeño general, con un Macro F1 de 0.8323 y los valores más altos de exhaustividad en ambas categorías, lo que evidencia un comportamiento más equilibrado y una mejor adaptación al dominio específico del corpus. Twitter-XLM-RoBERTa-base fue el segundo

mejor modelo, lo que sugiere que el preentrenamiento con un lenguaje cercano al de las redes sociales aporta ventajas, aunque inferiores a las de BETO en este dominio.

Modelo	Cat.	F1	Exh.	Prec.	Macro F1
TF-IDF-Reg. Log.	C1	0.7805	0.7901	0.7711	0.7377
	C3	0.6949	0.7736	0.6308	
DeBERTa-v3-base	C1	0.7544	0.7777	0.7325	0.7105
	C3	0.6666	0.7924	0.5394	
Twitter-XLM-R	C1	0.8048	0.8148	0.7951	0.7837
	C3	0.7627	0.8490	0.6923	
BETO	C1	0.8402	0.8765	0.8068	0.8323
	C3	0.8245	0.8867	0.7704	

Tabla 4: Desempeño comparativo de los modelos evaluados.

5.2 Discusión de los resultados

En conjunto, los resultados confirman que la detección de afirmaciones científicas constituye una tarea viable dentro de un corpus de *fact-checking* en español, siempre que se disponga de una capa de anotación suficientemente controlada y de modelos ajustados al idioma. El rendimiento alcanzado por BETO respalda la conveniencia de emplear modelos monolingües en escenarios en los que intervienen variaciones léxicas, estructuras propias del español y referencias semánticas sutiles asociadas al discurso científico.

Asimismo, la ausencia de instancias positivas suficientes en la categoría *referencia científica* constituye un hallazgo relevante del recurso. Lejos de ser únicamente una limitación experimental, este resultado refleja una característica del propio dominio: en las publicaciones analizadas es poco frecuente encontrar enlaces directos, DOI o citas formales a estudios científicos. En otras palabras, el contenido alude a la ciencia con mayor frecuencia de forma indirecta o narrativa, y no mediante referencias bibliográficas explícitas. Este fenómeno tiene implicaciones importantes para el diseño de herramientas de verificación, ya que obliga a combinar la clasificación textual con estrategias adicionales de recuperación de evidencia.

5.3 Análisis de errores

Aunque el mejor rendimiento global fue obtenido por BETO, la tarea sigue presentando dificultades semánticas, especialmente cuando el contenido no expresa de forma explícita una afirmación científica completa. En esta versión resumida, el análisis de errores se presenta de forma cualitativa, mediante la descripción de patrones representativos de confusión entre *afirmación científica verificable* y *contexto de investigación*, sin incluir una tabla detallada caso por caso.

Los errores más probables se concentran en tres situaciones: publicaciones que emplean léxico científico sin formular una proposición verificable completa, afirmaciones implícitas o incompletas que exigen inferencia contextual, y casos amplios de *contexto de investigación*, cuya variabilidad semántica dificulta una delimitación más estable. A ello se suma la escasa presencia de instancias asociadas a *referencia científica*, lo que limita tanto el entrenamiento de clasificadores como la recuperación posterior de evidencia académica. En conjunto, estos patrones refuerzan la idea de que la detección de discurso científico y la recuperación de evidencia deben entenderse como componentes complementarios dentro de un flujo híbrido de verificación asistida.

5.4 Resultados de la recuperación de evidencia académica

Como complemento del componente de detección, se analizó la viabilidad de recuperar literatura académica a partir de las afirmaciones científicas anotadas. Para ello, se distinguió entre afirmaciones científicas generales y afirmaciones documentalmente trazables, entendidas estas últimas como publicaciones cuya formulación alude a un único estudio o a una fuente documental suficientemente delimitable como para definir un objetivo de búsqueda específico. Sobre las 419 publicaciones confirmadas se aplicó este criterio de unicidad de estudio, identificándose 45 registros con condiciones adecuadas para una recuperación algorítmica dirigida.

Sobre este subconjunto se ejecutó el módulo de consulta a Crossref y PubMed, lo que generó un conjunto de búsqueda de 304 documentos candidatos. Este resultado sugiere que la principal restricción de la fase de recuperación no radica en la disponibilidad de las APIs utilizadas, sino en la baja especificidad documental de una parte importante del corpus. En efecto, muchas publicaciones contienen afirmaciones científicas verificables, pero no mencionan un único estudio ni aportan anclajes suficientes —como el nombre de la institución, la revista, un hallazgo distintivo o una referencia documental explícita— que permitan definir un objetivo de búsqueda inequívoco. En este sentido, los 45 casos recuperables corresponden a aquellas afirma-

ciones cuya formulación sí conserva huellas documentales suficientes para una recuperación algorítmica razonable, mientras que el resto requiere una mayor intervención interpretativa por parte del verificador humano.

5.5 Implicaciones para el fact-checking en español

Los resultados obtenidos sugieren que el recurso propuesto puede servir de base para futuras tareas de *fact-checking* científico en español. Primero, se demuestra que es posible especializar un corpus ecuatoriano de verificación mediante anotación manual orientada a un fenómeno discursivo concreto, preservando, además, la trazabilidad con el ecosistema de fact-checking del que procede. Además, los resultados confirman que los modelos monolingües para el español pueden alcanzar un rendimiento competitivo en la detección de este tipo de contenido, lo cual resulta especialmente relevante en escenarios en los que la verificación depende de matices léxicos y discursivos propios del idioma. Finalmente, evidencia que la recuperación de la evidencia académica no debe concebirse como una fase universal aplicable a toda afirmación científica, sino como un apoyo complementario para aquellos casos que conservan suficiente especificidad documental. En conjunto, estos hallazgos refuerzan la viabilidad de flujos híbridos en los que la clasificación automática actúe como filtro inicial y la recuperación de evidencia actúe como apoyo dirigido al trabajo del verificador humano.

5.6 Limitaciones del estudio

El trabajo presenta varias limitaciones. Por un lado, el recurso especializado se obtuvo a partir de un corpus de fact-checking ya existente, por lo que su distribución temática y discursiva depende de las fuentes previamente verificadas por el sistema y no abarca todo el espectro del discurso científico en redes sociales. Por otro lado, la categoría *referencia científica* no presentó suficientes instancias positivas para sostener una evaluación cuantitativa comparable a las otras dimensiones, lo que restringió la comparación principal a *afirmación científica verificable* y *contexto de investigación*.

Asimismo, la recuperación automática de evidencia académica solo fue viable para una parte reducida del subconjunto anotado, debido principalmente a la baja especificidad documental de muchas publicaciones. Finalmente, aunque la especificidad del recurso constituye una fortaleza, también limita la generalización inmediata de los resultados a otros contextos hispanohablantes o a fuentes con características discursivas distintas. Del mismo modo, aunque el proceso de anotación se apoyó en una guía explícita, una revisión contextual y una auditoría muestral del filtrado, una validación más formal de la confiabilidad interanotador constituye una línea de fortalecimiento metodológico para futuras versiones del recurso.

6 Conclusiones y trabajo futuro

Este trabajo presentó FactCheckEC-Sci, un recurso orientado a la detección de afirmaciones científicas en un corpus ecuatoriano de *fact-checking* en español. A partir de 6299 registros procedentes de fuentes de verificación, se desarrolló un proceso de filtrado preliminar y revisión manual que permitió consolidar un subconjunto especializado de 419 publicaciones con contenido científico, anotado en tres categorías no excluyentes: *afirmación científica verificable*, *referencia científica* y *contexto de investigación*.

Los resultados experimentales evidenciaron que la detección automática de afirmaciones científicas es una tarea viable en este dominio y que BETO obtuvo el mejor rendimiento global entre los modelos evaluados. Asimismo, la exploración de la recuperación automática de literatura académica mostró que esta línea puede apoyar flujos híbridos de verificación, aunque solo en los casos que conservan una especificidad documental suficiente. Desde esta perspectiva, FactCheckEC-Sci puede entenderse como una base especializada para estudiar cómo el discurso científico circula en entornos de verificación reales y cómo puede vincularse, de forma parcial pero útil, con evidencia académica recuperable. En conjunto, estos hallazgos refuerzan la viabilidad de flujos híbridos en los que la clasificación automática actúe como filtro inicial y la recuperación de evidencia actúe como apoyo dirigido al trabajo del verificador humano.

Como trabajo futuro, se plantea ampliar el subconjunto anotado, fortalecer la categoría *referencia científica*, explorar estrategias más avanzadas de recuperación y reranqueo de evidencia, y comparar BETO con estrategias *zero-shot* y *few-shot* en modelos de lenguaje grandes.

Bibliografía

Aloy-Mayo, M. y P. Rosso. 2026. Disinformation and conspiracy detection in spanish: A survey. *Procesamiento del Lenguaje Natural*, 76:65–78.

Barbieri, F., L. Espinosa Anke, y J. Camacho-Collados. 2022. XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond. En *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, páginas 258–266.

Bonet-Jover, A., R. Sepúlveda-Torres, y P. Martínez-Barco. 2023. Annotating reliability to enhance disinformation detection: Annotation scheme, resource and evaluation. *Procesamiento del Lenguaje Natural*, 70.

Cañete, J., G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, y J. Pérez. 2023. Spanish Pre-trained BERT Model and Evaluation Data.

CheckThat! Lab. 2025. CheckThat! at CLEF 2025 – Task 4: Scientific Web Discourse Processing. Available online: `https://checkthat.gitlab.io/clef2025/task4/` (accessed 2026-03-23).

Cochran, W. G. 1977. *Sampling Techniques*. John Wiley & Sons, New York, 3 edición.

Crossref. 2026. Crossref REST API. Available online: `https://www.crossref.org/documentation/retrieve-metadata/rest-api/` (accessed 2026-03-23).

Gómez-Adorno, H., J.-P. Posadas-Durán, F. Sánchez-Vega, y C. Porto Capetillo. 2021. Overview of FakeDeS at IberLEF 2021: Fake news detection in Spanish shared task. *Procesamiento del Lenguaje Natural*, 67:223–231.

Hafid, S., Y. S. Kartal, S. Schellhammer, K. Boland, D. Dimitrov, S. Bringay, K. Todorov, y S. Dietze. 2025. Overview of the CLEF-2025 CheckThat! Lab Task 4 on Scientific Web Discourse Processing. En *CLEF 2025 Working Notes*, volumen 4038 de *CEUR Workshop Proceedings*, páginas 722–732.

Hafid, S., S. Schellhammer, S. Bringay, K. Todorov, y S. Dietze. 2022. SciTweets: A dataset and annotation framework for detecting Scientific Online Discourse. En *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, páginas 3707–3717.

He, P., J. Gao, y W. Chen. 2023. DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing. En *International Conference on Learning Representations*.

National Center for Biotechnology Information. 2022. Entrez Programming Utilities (E-utilities) Help. NCBI Bookshelf. `https://www.ncbi.nlm.nih.gov/books/NBK25501/` (accessed 2026-03-23).

Reimers, N. y I. Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. En *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, páginas 3982–3992.

Toapanta Bernabe, M., M. A. Garcia-Cumbreras, y L. A. Urena-Lopez. 2024. Fake news detection and fact checking in X posts from Ecuador Chequea and Ecuador Verifica using Spanish Language Models. *Revista Tecnológica - ESPOL*, 36(2):158–173.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio para la Transformación Digital y de la Función Pública y el Plan de Recuperación, Transformación y Resiliencia —financiado por la UE–NextGenerationEU—, en el marco del proyecto "Desarrollo de Modelos ALIA". Asimismo, ha recibido financiación de los proyectos CONSENSO (PID2021-122263OB-C21), financiado por el MCIN/AEI/10.13039/501100011033 y la Unión Europea NextGenerationEU/PRTR; ROMANET (CERV-2024-CHAR-LITI-101215052), financiado por la Unión Europea mediante el programa "Ciudadanos, Igualdad, Derechos y Valores"; HEART-NLP-UJA (PID2024-156263OB-C21); VERITAS-H (AIA2025-163322-C64), financiado por el MICIU/AEI/10.13039/501100011033 y el FEDER/UE; y GALENO-IA (DGP_PIDI_2024_00852), financiado por la Unión Europea mediante S4Andalucía 2021–2027.

Esta investigación forma parte de la propuesta aprobada en la Convocatoria FCI 2023 de la Universidad de Guayaquil mediante Resolución N.º R-CSU-UG-SE34-313-14-09-2023 del Consejo Superior Universitario.