



Guía

La IA en medios de fact-checking: la experiencia de Newtral

Realizado por:  

En alianza con:  



Financiado por
la Unión Europea

may
20
26

* Introducción

Juan Sebastián Lozada, periodista de Colombiacheck, viajó a España en el marco de los intercambios de LatamChequea, la red de chequeadores latinoamericanos, y visitó las oficinas de Newtral con el objetivo de entender mejor cómo utilizan la inteligencia artificial y qué aprendizajes pueden aplicar otros chequeadores iberoamericanos.

En esta guía se explican distintas herramientas de IA desarrolladas y utilizadas en Newtral, se reflexiona sobre el uso de la IA en redacciones periodísticas y, basándose en las prácticas observadas, se dan recomendaciones puntuales para medios que están empezando a integrar IA en sus procesos, así como se listan distintas herramientas de inteligencia artificial que podrían ser útiles.

Autor: Juan Sebastián Lozada (Colombiacheck)

Editora: Verónica García López (Newtral)

Diseño: Matías Severo (Chequeado)

Realizado por:



Newtral

En alianza con:

LATAM CHEQUEA



Financiado por
la Unión Europea

*Este texto se ha elaborado en el marco del programa de intercambios de LatamChequea, la red de chequeadores latinoamericanos. Este intercambio formó parte del proyecto "Promover información fiable y combatir la desinformación en América Latina", coordinado por Chequeado y financiado por la Unión Europea. Su contenido es responsabilidad exclusiva de Colombiacheck y Newtral, y no refleja necesariamente los puntos de vista de la Unión Europea.

Índice

1. Introducción	4	Ir a página
2. Herramientas de IA desarrolladas y utilizadas en Newtral	6	Ir a página
› ClaimHunter y el editor de vídeo: automatización de la detección en fact-checking político	6	Ir a página
› FactFlow AI: monitoreo estructurado de desinformación en Telegram	8	Ir a página
› The Check: chatbot de verificación	11	Ir a página
3. Reflexiones sobre el uso de la IA: una mirada crítica	12	Ir a página
4. Aprendizajes y buenas prácticas observadas en Newtral	15	Ir a página
5. Recomendaciones para medios que están empezando a integrar IA en sus procesos	19	Ir a página
6. Mantener la calidad con humanos en el centro	30	Ir a página
7. Otras herramientas de IA de acceso gratuito o abierto útiles para fact-checking	32	Ir a página

1 Introducción

La inteligencia artificial (IA) ya forma parte de los procesos de muchas redacciones. En el ámbito periodístico, herramientas basadas en IA se utilizan para transcripción automática, monitoreo de redes sociales, análisis de grandes volúmenes de datos y apoyo en tareas de redacción o resumen. El [Reuters Institute Digital News Report 2024](#) señala que un número creciente de organizaciones de noticias está experimentando con IA generativa en distintas fases del trabajo informativo, aunque con importantes reservas respecto a la confianza del público y los estándares editoriales.

En la verificación de datos la presión necesaria es doble. Por un lado, el volumen de declaraciones públicas y contenidos virales ha aumentado de forma sostenida. De manera consistente, plataformas como [TikTok, Instagram y Reddit han ampliado su base de usuarios en los últimos años](#), lo que se correlaciona con un incremento en la cantidad de contenidos compartidos y difundidos en redes. Al mismo tiempo, el [análisis global de tendencias de desinformación](#) señala que los algoritmos de recomendación amplifican contenidos potencialmente desinformativos alimentados por determinados sesgos de los usuarios que contribuyen a su circulación viral.

Ante este escenario, algunas organizaciones de verificación han desarrollado herramientas propias para automatizar tareas preliminares relevantes en el proceso de fact-checking. Esta guía sintetiza la experiencia de la redacción de Newtral, uno de los medios de verificación de la Red Latam Chequea, que se ha planteado automatizar algunas fases del proceso de verificación, y, en el camino, aprender a partir de la prueba y el error, para luego compartir sus experiencias.

Esta redacción desarrolló modelos específicos para detectar afirmaciones verificables (“Claim Detection”) e identificar coincidencias con verificaciones previas (“Claim Matching”). Estas herramientas buscan reducir el tiempo dedicado a identificar posibles afirmaciones dentro de discursos extensos. Sin embargo, incluso en procesos como la transcripción automática, la redacción mantiene la revisión humana obligatoria debido a posibles errores en cifras o contextos complejos.

Integrar IA en la verificación no es solo una cuestión técnica. Implica tomar decisiones editoriales, éticas y metodológicas. ¿Dónde debe intervenir la automatización? ¿Qué riesgos introduce? ¿Qué partes del proceso no pueden o no deben delegarse? Las siguientes reflexiones parten de la experiencia práctica de una redacción que ha incorporado herramientas propias, pero que ha mantenido deliberadamente un enfoque de humanos en el centro del proceso (*human in the loop*).

La pregunta, entonces, ya no es si los medios de verificación utilizarán inteligencia artificial, sino cómo hacerlo sin comprometer calidad, transparencia y rendición de cuentas. Esta guía pretende explorar una respuesta para esa necesidad. Su objetivo es establecer criterios, límites y buenas prácticas para integrar herramientas de IA en procesos de verificación, manteniendo un enfoque de supervisión humana en todas las etapas críticas del trabajo editorial.

Herramientas de IA desarrolladas y utilizadas en Newtral

La integración de inteligencia artificial en Newtral responde a una estrategia de intervención en puntos específicos del proceso de fact-checking: detección inicial de afirmaciones, monitoreo estructurado de desinformación y acceso conversacional a verificaciones ya publicadas.

Desde 2018, la organización ha desarrollado herramientas propias que operan en distintas fases del flujo editorial. Estas son algunas de ellas:

ClaimHunter y el editor de vídeo: automatización de la detección en fact-checking político

Uno de los mayores desafíos en la verificación del discurso político es el tiempo invertido en escuchar intervenciones completas de los políticos o cargos públicos, transcribirlas e identificar qué frases contienen afirmaciones factuales susceptibles de verificación.

Para abordar este problema y optimizar el tiempo de trabajo de los fact-checkers, Newtral desarrolló un editor de vídeo basado en inteligencia artificial que automatiza parte de esta primera fase.

¿Cómo funciona?

El flujo operativo es el siguiente:

1. El equipo editorial define qué actos políticos hay en la agenda del día.
2. El departamento de documentación edita los discursos, intervenciones y otros.
3. El vídeo se sube al sistema.
4. El sistema genera una transcripción automática.

5. El modelo analiza la transcripción y marca las frases potencialmente verificables.
6. Los periodistas reciben una notificación en Slack cuando el contenido está disponible.

Cuando el verificador accede al editor, encuentra:

- El vídeo completo.
- La transcripción sincronizada.
- Las frases resaltadas.
- Un porcentaje que indica el nivel de verificabilidad.

The screenshot displays the ClaimHunter verification tool interface. On the left, there is a video player with a transcript overlay. The transcript is divided into segments (S1, S2) with corresponding verification percentages. A sidebar on the right lists 'Frases marcadas (45)' with their respective verification scores: 75%, 99%, 97%, and 97%. The video player shows a blue banner for 'Acción de Gobierno'.

Captura de pantalla de la herramienta de ClaimHunter desarrollada por Newtral.

Según Newtral, esta automatización ha permitido reducir en aproximadamente un 72% el tiempo empleado en la detección de afirmaciones factuales, que es la primera fase del proceso de verificación.

¿Qué hace el periodista?

La intervención humana ocurre en varios niveles:

- Revisa si la IA dejó fuera de la selección afirmaciones relevantes.
- Corrige frases mal clasificadas, y también confirma aquellas que son verificables.
- Evalúa el contexto en el que se produjo la afirmación.

Una periodista de la redacción explica que, cuando el periodista elimina frases incorrectas o añade otras no detectadas, el sistema aprende de esa interacción. Sin embargo, el reentrenamiento no es automático e indiscriminado: una persona del equipo de desarrollo señala que las anotaciones deben revisarse porque datos inconsistentes pueden degradar el modelo.

FactFlow AI: monitoreo estructurado de desinformación en Telegram

FactFlow AI fue desarrollada para abordar un problema específico: la dificultad de monitorear de forma sistemática la circulación de desinformación en Telegram, una plataforma donde los contenidos pueden viralizarse rápidamente y donde el rastreo manual resulta limitado.

La herramienta, presentada en el marco del JournalismAI Innovation Challenge, emplea modelos de inteligencia artificial para analizar texto, audio, vídeo e imágenes y detectar patrones narrativos asociados a desinformación.

¿Dónde interviene el criterio humano?

A diferencia de otras herramientas que actúan sobre un conjunto de datos ya delimitado, Fact Flow no opera sobre el universo completo de Telegram. Su funcionamiento depende de una decisión editorial previa.

El proceso comienza con una selección humana de cuentas o canales que el equipo identifica como nodos relevantes de circulación de desinformación. Esa primera “ingesta” no es automática: responde a criterios periodísticos basados en experiencia, seguimiento previo y conocimiento del ecosistema.

A partir de esa base, el modelo:

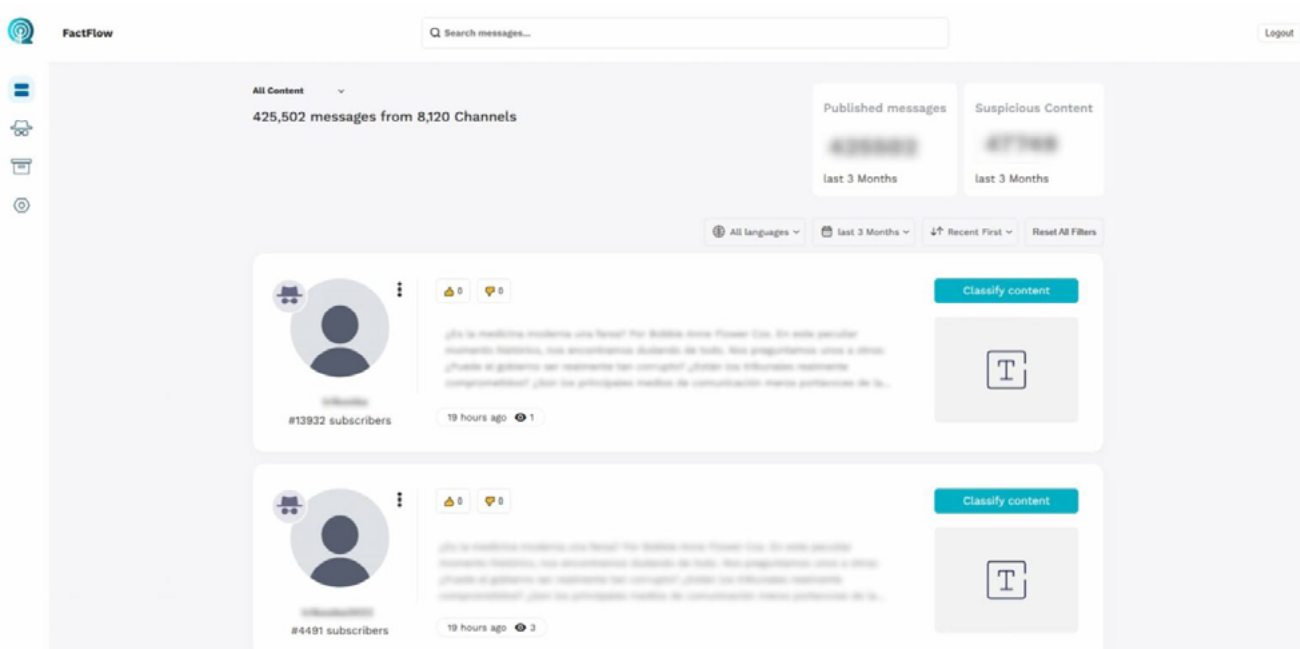
- Analiza los contenidos publicados.
- Identifica patrones narrativos.
- Detecta cuentas relacionadas.
- Amplía progresivamente su campo de rastreo.

Es decir, la IA extiende el monitoreo, pero lo hace a partir de un marco definido por periodistas.

Un periodista de la redacción dedicado a investigaciones de narrativas de desinformación señala que ninguna herramienta puede ofrecer una radiografía completa del ecosistema desinformativo; el monitoreo depende de cómo se estructuran los círculos o clusters iniciales y de qué se decide observar.

La herramienta no decide qué es un “nodo de desinformación”. Esa delimitación es resultado de decisiones editoriales que determinan:

- Qué comunidades observar.
- Qué narrativas priorizar.
- Qué idiomas o contextos geográficos son relevantes.



Captura de FactFlow, la herramienta de IA de Newtral para detectar bulos en Telegram.

¿Qué automatiza entonces FactFlow?

Una vez definida la arquitectura editorial inicial, la herramienta:

- Reduce el tiempo de rastreo manual.
- Detecta tendencias emergentes en tiempo real.
- Agrupa contenidos similares.
- Señala picos de viralidad.

Según datos públicos del proyecto, se han analizado más de 10 millones de mensajes y se han identificado más de 125.000 con desinformación potencial, en varios idiomas.

The Check: chatbot de verificación

The Check representa una línea distinta de desarrollo: el uso de modelos de lenguaje generativos para facilitar el acceso público a verificaciones ya realizadas. No genera nuevas verificaciones ni reemplaza la labor periodística. Su función es ampliar el acceso público a información ya contrastada y escalar la consulta.

El chatbot fue desarrollado en colaboración con otras organizaciones europeas y con apoyo del European Media Information Fund (EMIF). Su arquitectura se basa en un sistema de Generación Aumentada por Recuperación (RAG):

- El modelo no genera respuestas libres sin base documental.
- Recupera información de una base estructurada de verificaciones.
- Genera una respuesta natural respaldada por fuentes.
- Cita explícitamente esas fuentes, lo cual lo diferencia de muchos sistemas generativos convencionales.

Aprendizajes del desarrollo

El equipo enfrentó varios desafíos:

- Vacíos de información: si no existe una verificación previa sobre una afirmación concreta, el chatbot no puede inventarla.
- Calidad de las fuentes: en algunos casos, la respuesta era correcta pero las fuentes citadas no eran las más relevantes.
- Detección de idioma: fue necesario mejorar el filtrado lingüístico para evitar respuestas en idiomas distintos al de la consulta.

Reflexiones sobre el uso de la IA: una mirada crítica

Los verificadores revisaron y anotaron cientos de respuestas generadas por el sistema para mejorar precisión, claridad y atribución.

La primera tentación al hablar de inteligencia artificial en verificación es pensar en eficiencia. Las herramientas permiten procesar grandes volúmenes de información en tiempos muy reducidos. Como afirma uno de los miembros de la redacción de Newtral, es pasar de “arcos y flechas a un misil nuclear”. Esa potencia tecnológica amplía capacidades, pero no redefine por sí sola todo el proceso de verificación.

Por esa promesa de eficiencia, la integración de la IA en Newtral se ha centrado en los puntos donde el flujo de trabajo presenta los mayores cuellos de botella: las fases preliminares. Herramientas como *Claim Hunter* permiten transcribir declaraciones políticas, segmentar frases y marcar aquellas que podrían ser verificables. El objetivo no es decidir qué es verdadero o falso, sino reducir el tiempo invertido en localizar posibles afirmaciones verificables dentro de discursos extensos.

Aquí aparece el primer límite estructural: detectar no es decidir. La experiencia de Newtral muestra que la detectabilidad técnica de una frase no equivale a su relevancia pública. Una herramienta puede considerar verificables dos afirmaciones con la misma probabilidad estadística: una sobre una catástrofe con cientos de víctimas y otra sobre un porcentaje en una partida presupuestaria menor. Desde el punto de vista lingüístico, ambas afirmaciones son enunciados factuales. Desde el punto de vista editorial, no tienen el mismo peso. Este es el punto donde la eficiencia tecnológica encuentra su frontera. La diferencia entre ambas afirmaciones no es formal, sino contextual: impac-

to social, daño potencial, urgencia política y pertinencia pública. Ese filtro no está contenido en el algoritmo y aún pertenece al juicio periodístico.

Si la fase de detección se automatiza sin un criterio humano posterior, la agenda editorial podría desplazarse hacia lo que es técnicamente identificable y no necesariamente hacia lo que es socialmente relevante. Desde la redacción se advierte que delegar en una inteligencia artificial la definición de qué es importante para el periodismo y para la sociedad resulta problemático. La IA puede reconocer patrones, pero no comprende el peso histórico de un conflicto, la sensibilidad social de un tema o la urgencia política de una declaración.

Esta distinción entre detección y decisión conecta con un segundo elemento vital en la lucha contra la desinformación: la experiencia profesional acumulada. Con los años, los verificadores desarrollan una capacidad casi inmediata para distinguir entre opiniones, interpretaciones y afirmaciones medibles, para evaluar la disponibilidad de fuentes y para estimar si la verificación es viable en un tiempo razonable. Esa pericia no se basa únicamente en reglas formales, sino en práctica sostenida y conocimiento del entorno informativo. Es una forma de criterio contextual que no se reduce a la clasificación estadística.

Por eso, incluso cuando la automatización interviene en etapas preliminares, no elimina la necesidad de supervisión constante. En los modelos de detección de afirmaciones (Claim Hunter), por ejemplo, el sistema mejora mediante retroalimentación: los periodistas pueden borrar frases mal clasificadas o añadir aquellas que el modelo no detectó, y esos datos se utilizan para reentrenarlo. Sin embargo, si las anotaciones son inconsistentes o responden a criterios divergentes, el sistema puede degradar su desempeño. La herramienta depende de la coherencia metodológica humana para mantenerse fiable. Demostrando que este no es un sistema autónomo, sino que requiere acompañamiento permanente.

La monitorización de desinformación ofrece otro ejemplo de esta interdependencia. No existe una herramienta que proporcione una radiografía completa de lo que circula en redes. El resultado depende de cómo el equipo organiza listas de cuentas, define burbujas ideológicas o establece palabras clave estratégicas. En otras palabras, la tecnología ejecuta una estrategia previamente diseñada por periodistas.

En la práctica, la redacción reconoce que no ha automatizado el proceso integral ni es su intención hacerlo. La verificación incluye decisiones narrativas: cómo explicar el resultado, qué formato utilizar, cómo contrarrestar la viralidad, que siguen siendo profundamente periódicas. Incluso la forma de combatir una desinformación viral (video, hilo en redes, contacto con la institución afectada) es una decisión editorial y estratégica que excede cualquier clasificación algorítmica.

De ahí que la posición general en la redacción de Newtral sea una de cautela. Existe una postura consistente: no rechazo a la IA, pero tampoco adopción acrítica. Se prueban herramientas, se evalúan resultados y se mantiene supervisión constante por parte de humanos. Esa cautela no es resistencia al cambio, más bien es coherente con el escepticismo profesional que define el trabajo de verificación.

En conjunto, la experiencia muestra una secuencia clara: la IA puede mejorar la eficiencia en la detección, la detección requiere priorización, la priorización exige criterio humano, y el criterio humano implica responsabilidad editorial. La tecnología multiplica la capacidad operativa, pero no redefine el núcleo del proceso. Mantener explícita esta cadena es fundamental para preservar la calidad y la credibilidad del trabajo de verificación.

Aprendizajes y buenas prácticas observadas en Newtral

Más allá de los límites conceptuales de la automatización, ya discutidos en la sección anterior, la experiencia de Newtral ofrece aprendizajes operativos y organizacionales que resultan especialmente relevantes para medios que comienzan a integrar herramientas de inteligencia artificial en sus procesos de verificación.

Estos aprendizajes no se centran únicamente en qué automatizar, sino en cómo hacerlo sin alterar la coherencia editorial, la organización interna y la responsabilidad pública.

1. La IA obliga a formalizar criterios que antes podían ser implícitos

Uno de los efectos menos visibles de la integración de IA exige gobernanza interna clara y criterios editoriales formalizados. En los sistemas de detección de afirmaciones, el modelo aprende a partir de anotaciones humanas. Cuando un periodista corrige una frase mal segmentada o añade una afirmación no detectada, está traduciendo su criterio editorial en datos estructurados.

Sobre ClaimHunter, la herramienta que usan en Newtral, una periodista acota que “aprende cuando borras frases que no son verificables o añades otras que sí lo son”. Cada decisión humana pasa a formar parte del entrenamiento del modelo. Esto introduce una exigencia metodológica nueva: si cada integrante del equipo aplica criterios distintos sobre qué constituye una afirmación verificable, el sistema no solo pierde precisión, sino coherencia.

La automatización, por tanto, no solo requiere supervisión, sino estandarización. La herramienta actúa como un espejo metodológico: obliga a la

redacción a acordar y documentar criterios comunes sobre qué es verificable, qué no lo es y bajo qué condiciones se clasifica una afirmación.

2. La automatización transforma la organización del trabajo

Otro aprendizaje relevante es que la IA no solo acelera procesos, también altera el ritmo y la distribución del trabajo dentro de la redacción. Integrar IA no es sólo incorporar una herramienta, sino redefinir dinámicas internas y cargas de trabajo. Al reducir el tiempo de detección y segmentación de frases, se redistribuyen los esfuerzos hacia fases posteriores: análisis contextual, consulta de fuentes, redacción explicativa y estrategia de difusión.

Esto implica una reorganización implícita del flujo de trabajo. El valor añadido se desplaza hacia la interpretación y la contextualización. Además, el hecho de que el sistema pueda detectar múltiples afirmaciones en poco tiempo puede generar un aumento en el volumen potencial de verificaciones, lo que exige mecanismos claros de priorización para evitar sobrecarga editorial.

3. La estrategia de monitoreo sigue siendo editorial

En el ámbito de la monitorización de desinformación, las herramientas permiten rastrear palabras clave, listas de cuentas o burbujas ideológicas, pero no determinan por sí mismas qué segmentos del ecosistema informativo deben observarse ni cómo deben organizarse.

Un miembro de la redacción apunta a que no existe una herramienta que ofrezca una “radiografía completa” de lo que circula en redes. Todo depende de cómo se construye la arquitectura de seguimiento: qué listas se crean, qué burbujas se definen, qué cuentas se priorizan.

Esto significa que la calidad del monitoreo no descansa principalmente en la sofisticación tecnológica, sino en la estrategia editorial que la sostiene. La herramienta ejecuta una lógica previamente definida por periodistas; no la sustituye ni la corrige.

4. La integración tecnológica requiere coordinación interdisciplinaria

En el caso de Newtral, el desarrollo y ajuste de herramientas propias fue un ejercicio interdisciplinar: un diálogo entre los desarrolladores y los periodistas. La implementación no fue únicamente una decisión editorial sino organizacional.

Como explica un miembro del equipo de desarrollo, el modelo no surge en abstracto, sino como respuesta a problemas concretos del flujo de trabajo. Esta coordinación resulta determinante para que la herramienta responda a necesidades reales y no a supuestos teóricos.

Para redacciones que no cuentan con equipos técnicos internos, este aprendizaje es especialmente relevante: la integración tecnológica exige diálogo continuo entre lógicas distintas.

5. La IA introduce un nuevo tipo de riesgo reputacional

Aunque los miembros de Newtral no ven la automatización como amenaza directa, sí muestran una actitud consistente de cautela. No se trata de desconfianza hacia la tecnología, sino de conciencia sobre el riesgo que implicaría depender acríticamente de sus resultados.

En un entorno donde la credibilidad es el principal activo del fact-checking, un error derivado de la confianza excesiva en una herramienta po-

dría tener consecuencias reputacionales significativas. Por ello, la supervisión constante no es solo una práctica metodológica; es una estrategia de protección institucional.

6. Delimitar explícitamente lo no automatizable

Finalmente, esta redacción resalta un punto fundamental: el proceso completo de verificación no está automatizado ni se pretende que lo esté. Las decisiones narrativas, como cómo explicar el resultado, qué formato utilizar, cómo responder a la viralidad, siguen siendo profundamente periodísticas.

Una periodista de la redacción lo planteaba con claridad: la herramienta puede asistir, pero el criterio y la responsabilidad siguen siendo humanos.

Este límite explícito evita la ilusión de automatización total y protege el núcleo profesional del fact-checking. La tecnología amplía capacidades operativas, pero no redefine la responsabilidad editorial ni el juicio contextual que sostienen el trabajo de verificación.

5 Recomendaciones para medios que están empezando a integrar IA en sus procesos

La incorporación de inteligencia artificial en medios de verificación no es únicamente una decisión tecnológica. Es una decisión organizacional, metodológica y estratégica. Para redacciones que están comenzando a explorar estas herramientas o que las han integrado recientemente, las siguientes recomendaciones buscan orientar una adopción gradual, sostenible y alineada con estándares editoriales.

1. Comenzar con un diagnóstico interno, no con una herramienta

El error más común al iniciar procesos de automatización es empezar por la solución tecnológica y no por el problema editorial. El desarrollo en Newtral no comenzó con una herramienta aislada, sino con la identificación de dos problemas concretos: la detección de frases verificables y el “claim matching”. Es decir, se partió de tareas específicas dentro del flujo de trabajo que podían ser formalizadas.

Antes de incorporar cualquier sistema, el medio debería preguntarse qué fase del proceso está generando mayor carga operativa. En el caso del fact-checking político, la necesidad era clara: escuchar entrevistas completas consumía un tiempo desproporcionado, e incluso mayor trabajo durante los fines de semana. La herramienta se diseñó para reducir esa fricción, no para reemplazar la verificación.

Sin un diagnóstico claro del cuello de botella, la herramienta puede convertirse en una capa adicional de trabajo, especialmente si no resuelve el problema real al que la redacción se enfrenta.

2. Establecer una política interna de uso antes de implementar

Muchos medios comienzan a utilizar herramientas de IA de manera informal, sin un marco interno claro. Esto genera inconsistencias y riesgos. Newtral elaboró una guía interna para fijar límites en el uso de IA generativa y evitar comprometer la metodología. Este antecedente muestra que la formalización temprana no es un trámite burocrático, sino un mecanismo de protección editorial.

Desarrollar una política interna implica responder de forma explícita a varias cuestiones operativas. En primer lugar, delimitar qué tareas pueden apoyarse en IA. En el caso analizado, las herramientas intervienen en fases preliminares como transcripción automática, detección de frases potencialmente verificables o búsqueda de coincidencias con verificaciones previas. No sustituyen la consulta de fuentes, la interpretación contextual ni la redacción final. Una política clara debería especificar ese alcance: la IA puede asistir en tareas repetitivas o sistematizables, pero no en decisiones editoriales sustantivas.

En segundo lugar, definir qué fases requieren revisión humana obligatoria. El periodista puede revisar errores incluso después de la transcripción automática, especialmente en cifras o nombres propios. Este tipo de verificación manual no es opcional: forma parte del control de calidad. Una política interna debería establecer que toda tarea automatizada sea validada por un humano antes de convertirse en insumo editorial.

También es necesario clarificar qué herramientas están autorizadas y bajo qué condiciones. La referencia a una guía interna sobre IA generativa muestra que no todas las aplicaciones son equivalentes ni pueden utilizarse indiscriminadamente. Esto incluye decidir si se permiten herramien-

tas externas, cómo se protegen los datos y qué usos están expresamente excluidos.

Otra dimensión clave es la documentación de los procesos automatizados. Si el modelo aprende de las anotaciones humanas, conviene registrar cómo se corrigen frases, cómo se clasifican afirmaciones y qué criterios se aplican. Esta trazabilidad no sólo mejora el sistema; también permite auditar decisiones en caso de discrepancias internas o cuestionamientos externos.

Finalmente, una política interna debe asignar responsabilidades claras. ¿Quién supervisa el funcionamiento del sistema? ¿Quién detecta patrones de error? ¿Quién decide cuándo reentrenar el modelo o ajustar parámetros? La redacción de Newtral comenta que debe haber interacción constante entre perfiles técnicos y editoriales, lo que sugiere que la supervisión no puede recaer exclusivamente en una sola área.

En conjunto, establecer una política antes de escalar el uso de IA permite evitar improvisaciones y asegura que la herramienta opere bajo reglas editoriales previamente acordadas, en lugar de redefinirlas de manera implícita.

3. Implementar pilotos controlados antes de escalar

La integración tecnológica debe hacerse en fases. En lugar de adoptar una herramienta en toda la redacción desde el inicio, es recomendable:

- Probarla en un equipo reducido.
- Medir resultados durante un período definido.
- Evaluar errores, sesgos y limitaciones.
- Ajustar procesos antes de ampliarla.

La experiencia técnica descrita por el equipo de Newtral muestra múltiples iteraciones antes de alcanzar estabilidad: problemas de transcripción con acentos latinoamericanos, errores en signos de interrogación y dificultades con ruido de fondo. Estos ajustes no se resolvieron en una sola fase. Hubo pruebas, entrenamiento con datos ruidosos y reentrenamientos progresivos. Esta dinámica confirma que la integración tecnológica requiere validación en contexto real antes de expandirse a toda la redacción.

Escalar sin fase piloto puede ocultar limitaciones que solo aparecen bajo condiciones cotidianas de trabajo.

4. Invertir en alfabetización tecnológica del equipo

Las herramientas implementadas pueden ser intuitivas en su uso, pero su funcionamiento interno es más complejo de lo que la interfaz sugiere. Que un sistema sea sencillo de operar no significa que sea sencillo de comprender.

La alfabetización tecnológica, en este contexto, implica que el equipo editorial entienda al menos cuatro dimensiones del sistema con el que trabaja.

Primero, ¿cómo está entrenado? En el caso analizado, el modelo evoluciona a partir de datos anotados por periodistas y ha pasado por distintas arquitecturas (desde reglas lingüísticas hasta modelos multilingües más avanzados). Conocer esta evolución permite comprender que el rendimiento no es estático ni universal, sino dependiente de datos y contexto.

Segundo, ¿qué variables afectan su desempeño? Pueden existir dificultades asociadas a acentos, ruido de fondo, cambios de idioma o estructuras sintácticas complejas. Esto significa que el sistema no opera con igual precisión en todas las condiciones. Entender estas limitaciones permite

interpretar sus resultados con mayor criterio.

Tercero, ¿cómo procesa la información? Por ejemplo, el sistema segmenta frases y clasifica la probabilidad de verificabilidad. Comprender que esa clasificación es probabilística y no determinista ayuda a evitar lecturas absolutas de la salida del modelo.

Cuarto, ¿cómo mejora o se degrada con el uso? El modelo aprende de anotaciones humanas y que la calidad de esos datos afecta directamente su rendimiento. Esto implica que la herramienta no es un sistema cerrado, sino un entorno dinámico que evoluciona según la interacción con el equipo.

Invertir en alfabetización tecnológica significa crear espacios de formación interna donde estos aspectos se expliquen de manera clara: cómo funciona el modelo, qué métricas utiliza, en qué contextos tiene menor precisión y cómo se interpretan sus resultados.

No se trata de convertir a los periodistas en desarrolladores, sino de asegurar que comprendan la lógica básica del sistema que forma parte de su flujo de trabajo. Sin esta comprensión, la herramienta puede ser utilizada correctamente en apariencia, pero malinterpretada en su alcance real.

En entornos donde la tecnología interviene en etapas relevantes del proceso editorial, entender el funcionamiento del sistema es parte de la competencia profesional contemporánea.

5. Prever mecanismos de auditoría y trazabilidad

Cuando se incorpora inteligencia artificial en fases del proceso editorial, la

cuestión no es solo qué hace la herramienta, sino cómo se puede reconstruir posteriormente el recorrido de una verificación. La trazabilidad no es un elemento técnico accesorio, sino una condición metodológica.

En el caso analizado, el flujo no se limita a que el sistema “detecte frases”, sino que esas frases quedan registradas en una base de datos y se integran en un circuito organizado de trabajo. Esto implica que las decisiones no desaparecen en el sistema, sino que forman parte de un proceso documentado.

Para medios que comienzan a integrar IA, prever mecanismos de auditoría significa poder responder con claridad a preguntas concretas:

- ¿Qué herramienta intervino en el proceso?
- ¿En qué momento del flujo editorial lo hizo?
- ¿Qué transformaciones produjo sobre el contenido original?
- ¿Qué decisiones editoriales se tomaron después?
- ¿Qué versión del modelo estaba activa en ese momento?

Estas preguntas no son hipotéticas. En contextos de fact-checking, donde la credibilidad y la transparencia son activos centrales, cualquier error o controversia puede requerir una reconstrucción detallada del proceso. Si no existen registros claros, la organización queda expuesta a incertidumbre interna y cuestionamientos externos.

Además, la trazabilidad no sólo protege frente a críticas públicas; también permite evaluar internamente el desempeño del sistema. Saber cuándo y cómo intervino la herramienta facilita identificar patrones de error, ajustar procesos o redefinir su alcance.

Diseñar mecanismos de auditoría desde el inicio evita que la automatización se vuelva opaca con el tiempo. Documentar el uso de IA, sin necesi-

dad de exponer detalles técnicos complejos, forma parte de una cultura organizacional que entiende la tecnología como parte del método, no como una capa invisible añadida al final.

En redacciones que trabajan con estándares de verificación, la trazabilidad no es opcional: es coherente con la lógica de transparencia que ya rige el trabajo periodístico.

6. Evaluar el impacto en la confianza pública

La preocupación por los límites de la automatización no es únicamente técnica, sino institucional. Cuando se plantea la posibilidad de automatizar más fases del proceso, la respuesta no es entusiasta ni defensiva, sino prudente.

Una periodista de la redacción afirma que el proceso de verificación no está automatizado por completo en la redacción de Newtral, ni se pretende que lo esté. Esta afirmación no se formula como una limitación tecnológica, sino como una decisión editorial. Es decir, existe una frontera deliberada.

Asimismo, se advierte sobre el riesgo de delegar decisiones clave, como qué temas priorizar o qué afirmaciones considerar relevantes, a sistemas automatizados. Este punto es especialmente significativo: la agenda informativa no es una variable técnica, sino una expresión de criterio editorial.

Estas posiciones reflejan una conciencia clara de que la credibilidad del medio no depende únicamente de la precisión factual, sino de la percepción pública de responsabilidad. Si la audiencia percibe que las decisiones centrales están mediadas por un sistema automatizado sin intervención humana visible, la legitimidad puede verse afectada, incluso si el resultado factual es correcto.

En este sentido, evaluar el impacto en la confianza pública implica tomar decisiones anticipadas sobre transparencia. ¿Se explicará qué papel tiene la IA en el proceso? ¿Se comunicará que interviene en fases preliminares, pero no en la calificación final? ¿Se hará explícita la responsabilidad humana en la decisión editorial?

La cautela frente a la automatización total no responde a desconfianza tecnológica, sino a la necesidad de preservar el núcleo profesional del fact-checking. La línea que se traza al no automatizar decisiones editoriales sustantivas, es también una línea de protección reputacional.

Para medios que comienzan a integrar IA, el desafío no es solo técnico, sino comunicativo: anticipar cómo el uso de estas herramientas será percibido y asegurar que la responsabilidad editorial siga siendo clara hacia los lectores.

7. Evitar dependencia excesiva de proveedores externos

Muchas herramientas de IA operan como servicios externos (APIs, plataformas SaaS o modelos alojados por terceros) cuya lógica de funcionamiento, actualizaciones y condiciones de uso no están bajo control del medio.

Aunque en el caso analizado existe desarrollo propio, hay una conciencia clara de que los modelos evolucionan constantemente y que su desempeño depende de decisiones técnicas que no siempre son visibles para el usuario final. El sistema ha pasado por distintas arquitecturas y entrenamientos, y el rendimiento varía según idioma y volumen de datos disponibles. Esta evolución interna ilustra un punto más amplio: cuando el modelo no es propio, esos cambios pueden producirse sin que el medio tenga

control directo sobre ellos.

Si una redacción depende exclusivamente de un proveedor externo, pueden surgir riesgos como:

- Cambios en políticas de uso o acceso.
- Incrementos de costos.
- Modificaciones en el modelo que alteren el rendimiento sin previo aviso.
- Restricciones sobre datos compartidos o almacenados.

Muchas soluciones tecnológicas no están pensadas específicamente para flujos periodísticos. Esto implica que, además de los riesgos contractuales o técnicos, existe una posible brecha entre las decisiones del proveedor y las necesidades del medio.

Para redacciones que comienzan a integrar IA, la recomendación no es evitar herramientas externas, sino evaluar cuidadosamente el grado de dependencia que se está creando. Esto incluye preguntarse:

- ¿Qué datos se están enviando a terceros?
- ¿Qué ocurre si el proveedor cambia condiciones o elimina una funcionalidad?
- ¿Existen alternativas o planes de contingencia?
- ¿El medio podría continuar operando si ese servicio deja de estar disponible?

La automatización puede aumentar la eficiencia, pero si no se evalúan estos factores, también puede introducir vulnerabilidad estructural. No depender exclusivamente de un proveedor sin una evaluación de riesgos permite preservar autonomía editorial y operativa.

8. Revisar periódicamente la necesidad real de la herramienta

La incorporación de IA no es irreversible ni definitiva. Toda herramienta tecnológica debería estar sujeta a evaluación periódica, especialmente en entornos editoriales donde los procesos y necesidades evolucionan con rapidez.

Integrar una solución en el flujo de trabajo puede generar la impresión de que la decisión está cerrada. Sin embargo, las condiciones cambian: aparecen nuevas alternativas, se modifican costos, surgen actualizaciones que alteran funcionalidades o el propio medio redefine sus prioridades estratégicas.

Por ello, es recomendable establecer momentos formales de revisión para preguntarse:

- ¿La herramienta sigue resolviendo el problema que motivó su adopción?
- ¿Realmente mejora procesos o solo los reorganiza?
- ¿El costo-beneficio sigue siendo razonable?
- ¿Existen alternativas más adecuadas o eficientes?
- ¿Ha generado dependencias no previstas?

Esta revisión no implica desconfianza hacia la tecnología, sino responsabilidad organizacional. En el ámbito del fact-checking, donde la metodología y la credibilidad son activos centrales, mantener herramientas por inercia puede afectar tanto la eficiencia como la coherencia editorial.

Tratar la integración tecnológica como un proceso dinámico permite adaptar la estrategia del medio a cambios técnicos, económicos o editoriales sin quedar atado a elecciones pasadas.

Integrar inteligencia artificial en medios de verificación no es una cuestión de modernización simbólica ni de alineación con tendencias tecnológicas. Es una decisión que afecta los flujos de trabajo, responsabilidades editoriales, gobernanza interna y confianza pública.

Los medios que comienzan este camino deberían asumir que la automatización no sustituye la metodología de verificación, sino que la tensiona y la obliga a formalizarse. Por ello, el éxito no depende únicamente de la herramienta elegida, sino de la claridad con la que se definan sus límites, su supervisión y su impacto.

Mantener la calidad con humanos en el centro

La integración de inteligencia artificial en procesos de verificación no es un problema tecnológico, sino un desafío de calidad editorial. En fact-checking, la credibilidad no depende únicamente del resultado final sino del método que conduce a esa conclusión. Cuando intervienen sistemas automatizados, preservar la calidad implica reforzar ciertos principios que sostienen el trabajo periodístico.

La calidad depende del método, no solo del resultado

Una verificación correcta obtenida mediante un proceso opaco o poco riguroso no fortalece la credibilidad. El uso de IA no puede reemplazar las etapas esenciales del contraste: identificación de fuentes, revisión documental, contextualización y explicación pública. La automatización puede asistir en fases preliminares, pero la estructura metodológica debe permanecer intacta.

Mantener a los humanos en el centro significa que el proceso editorial (cómo se investiga, cómo se contrasta y cómo se explica) no se redefine en función de la herramienta.

La interpretación es parte del estándar de calidad

En fact-checking, la calidad no se limita a comprobar si un dato aislado es correcto. Una afirmación puede contener cifras verdaderas y, sin embargo, estar presentada fuera de contexto, omitir información relevante o utilizar comparaciones que distorsionan su significado. Evaluar estos matices es parte del trabajo de verificación. No basta con comprobar datos, también es necesario analizar cómo se utilizan y contrastar con expertos en la materia si la interpretación es correcta y aplicable al caso concreto. Esto por el momento, solo lo puede hacer un humano.

Los sistemas automatizados pueden identificar patrones formales en el lenguaje, pero no evalúan intención, contexto ni efecto discursivo. Preservar la calidad implica asegurar que esa dimensión interpretativa no se delegue ni se simplifique en función de lo que la herramienta puede detectar.

La coherencia editorial es un activo crítico

Cuando un medio utiliza herramientas automatizadas, la consistencia interna se vuelve aún más importante. Si distintos miembros del equipo aplican criterios divergentes sobre clasificación o priorización, el resultado no solo afecta la herramienta, sino la percepción pública de coherencia.

Mantener a los humanos en el centro implica alinear criterios, revisar discrepancias y asegurar que la metodología sea estable a lo largo del tiempo. La calidad también es consistencia.

La supervisión es garantía de rigor

La intervención humana no debe limitarse a revisar errores evidentes como una palabra mal escrita en una transcripción. Debe consistir en evaluar críticamente la pertinencia de lo que el sistema sugiere y detectar posibles omisiones. En entornos automatizados, el rigor no se reduce más bien se desplaza hacia la supervisión consciente.

Mantener la calidad implica asumir que toda herramienta puede fallar y que el control editorial es una responsabilidad permanente.

En síntesis, el enfoque *human-in-the-loop* no es un principio abstracto. Es una forma de proteger la calidad metodológica, la coherencia editorial y la responsabilidad pública en un entorno donde la tecnología amplía capacidades, pero no sustituye criterio.

Otras herramientas de IA de acceso gratuito o abierto útiles para fact-checking

Detección y búsqueda de verificaciones previas

- [Google Fact Check Tools](#)

Permite buscar verificaciones ya realizadas por organizaciones adheridas a estándares internacionales. Es útil para identificar rápidamente si una afirmación ya fue analizada, comparar enfoques y evitar duplicación de trabajo.

Transcripción automática de audio y video

- [Whisper \(open-source\)](#) y [Otter.ai \(plan gratuito\)](#)

Permiten convertir audio o video en texto de manera automática. Resultan útiles para procesar ruedas de prensa, discursos políticos o contenidos virales extensos. Pueden cometer errores en cifras, nombres propios o acentos; requieren revisión humana posterior.

Verificación de imágenes y video

- [InVID](#) (extensión de navegador)

- Permite descomponer videos en fotogramas clave, realizar búsquedas inversas y analizar metadatos disponibles. Estas herramientas pueden no detectar todas las manipulaciones y deben combinarse con contraste adicional.

- [Bellingcat TikTok Date Extractor](#)

- [Watch Frame by Frame](#)

- [Filmot](#)

Búsqueda académica y contextualización

- [Semantic Scholar](#) y [OpenAlex](#)

Plataformas que incorporan capacidades de búsqueda semántica sobre literatura científica. Ayudan a localizar estudios relevantes de forma rápida para afirmaciones técnicas o científicas.

Archivo y trazabilidad

- [Wayback Machine](#)

Permite consultar versiones históricas de páginas web. Puede usarse para documentar cambios en contenidos digitales y conservar evidencia. Sin embargo, no todas las páginas están archivadas de manera sistemática.

- [Perma](#)
- [MediaVault](#)

Identificación de contenido sintético

- [Hive Moderation](#)
- [Resemble AI](#)
- [ElevenLabs](#)
- [SynthID de Google](#)

Guía

La IA en medios de fact-checking: la experiencia de Newtral

Realizado por:  

En alianza con:  

 Financiado por
la Unión Europea

may
20
26

LATAM  CHEQUEA

in 