



Bots en el periodismo: patrones discursivos de noticias creadas por Inteligencia Artificial y noticias escritas por humanos

Bots in journalism: discursive patterns of news created by Artificial Intelligence and news written by humans

Bots no jornalismo: padrões discursivos de notícias criadas por Inteligência Artificial e notícias escritas por humanos

Victoria Daniela Rodríguez-Flores¹
Universidad Autónoma de Chihuahua
victoria_drf@hotmail.com

David Picazo²
Universidad Autónoma de Chihuahua
ipicazo@uach.mx

Claudia Patricia Contreras³
Universidad Autónoma de Chihuahua
ccontreras@uach.mx

¹ Licenciada en Lengua Inglesa por la Universidad Autónoma de Chihuahua. Graduada de la Maestría en Periodismo y Poder. Profesora de la Universidad Tecnológica de Chihuahua, Unidad Académica BIS. ORCID: <https://orcid.org/0009-0009-1404-3769>

² Doctor en Educación, profesor de tiempo completo ATC de la Universidad Autónoma de Chihuahua. Miembro del Sistema Nacional de Investigadores. Profesor en licenciatura, maestría y doctorado, con materias relacionadas con investigación, estadística y sistemas de información. Línea de investigación: Análisis del discurso académico y científico. Ha escrito más de 20 artículos científicos y 5 capítulos de libro. ORCID: <https://orcid.org/0000-0003-0521-6954>

³ Doctora en Periodismo, profesora de tiempo completo ATC de la Universidad Autónoma de Chihuahua. Miembro del Sistema Nacional de Investigadores. Profesora de licenciatura y maestría con materias relacionadas con investigación, lingüística y análisis del discurso. Línea de investigación: Análisis del discurso académico y científico. Ha escrito artículos científicos y capítulos de libros. ORCID: <https://orcid.org/0000-0003-4162-3055>

Fecha de recepción: 5 de marzo de 2025

Fecha de aprobación: 10 de mayo de 2025

Fecha de publicación: 2 de junio de 2025

Resumen

El periodismo deportivo está empleando con mayor frecuencia la inteligencia artificial (IA) para automatizar la generación de notas en eventos de gran cobertura, sin embargo, es preciso reconocer que existen diferencias discursivas entre las noticias escritas por IA y aquellas escritas por humanos. El objetivo de este trabajo fue identificar patrones discursivos en los tuits publicados en X (antes Twitter) generados por la inteligencia artificial denominada OlyBot y por periodistas deportivos en los Juegos Olímpicos de Río de Janeiro 2016 y Pyeongchang 2018. Se aplicó el índice de niebla de Gunning para analizar una muestra de 30 tuits emitidos por el sistema automatizado y 30 tuits emitidos por periodistas, en cada uno de los eventos. Los resultados indican que la redacción de esta IA es simple y concisa, pues hace uso de únicamente una oración por tuit. Las partes del discurso encontradas fueron: sustantivos, verbos y adjetivos. En Río de Janeiro 2016, ambos, periodistas (10.54) y bot (10.44), emitieron notas que se pueden leer sin dificultad, según el Índice de niebla de Gunning. El resultado más relevador se encontró en el análisis de legibilidad de las notas emitidas en Pyeongchang 2018, donde los periodistas alcanzaron un valor de 14.63 contra 10.8 de OlyBot, lo que indica, según el modelo utilizado, que se trata de textos difíciles de leer, que no son de consumo popular. Con el aumento de palabras, el bot de inteligencia artificial mantuvo su nivel de legibilidad en ambos eventos, mientras que el periodista incrementó la complejidad de sus notas, haciéndolas más ininteligibles. Las implicaciones de estas reflexiones son de interés para tecnólogos, periodistas y lingüistas, así como para la sociedad en general, pues se presenta el reto de discernir y valorar aquellos textos que actualmente pueden ser creados por autómatas o por periodistas, sin menoscabo de la legibilidad de su lectura.

Palabras clave: inteligencia artificial, periodismo automatizado, análisis del discurso, redes sociales, bots

Abstract

Sports journalism is increasingly employing artificial intelligence (AI) to automate the generation of notes in events of great coverage. However, it is necessary to recognize that there are discursive differences between the news written by AI and those written by humans. The objective of this work was to identify discursive patterns in the tweets generated by the artificial intelligence called OlyBot and by sports journalists in the Olympic Games of Rio de Janeiro 2016 and Pyeongchang 2018. Gunning's fog index was applied to analyze a sample of 30 tweets issued by the automated system and 30 tweets issued by journalists, in each event. The results indicate that the wording of this AI is simple and concise, as it uses only one sentence per tweet. The parts of speech found

were: nouns, verbs, and adjectives. In Rio de Janeiro 2016, both journalists (10.54) and bot (10.44) issued notes that can be read without difficulty, according to the Gunning Fog Index. The most relevant result was found in the readability analysis of the notes issued in Pyeongchang 2018, where journalists reached a value of 14.63 against OlyBot's 10.8, indicating, according to the model used, that these are difficult-to-read texts, which are not for popular consumption. With the increase in words, the artificial intelligence bot maintained its readability level in both events, while the journalist increased the complexity of his notes, making them more unintelligible. The implications of these reflections are of interest to technologists, journalists, and linguists, as well as to society in general since they present the challenge of discerning and assessing those texts that can currently be created by automatons or by journalists, without undermining their readability.

Keywords: artificial intelligence, automated journalism, discourse analysis, social networks, bots

Resumo

O jornalismo esportivo está cada vez mais utilizando inteligência artificial (IA) para automatizar a geração de notas em eventos de alta cobertura jornalística, no entanto, é preciso reconhecer que há diferenças no estilo entre notícias escritas por IA e aquelas escritas por humanos. O objetivo deste trabalho foi identificar padrões de estilo em tweets gerados pela inteligência artificial chamada OlyBot e por jornalistas esportivos nos Jogos Olímpicos do Rio de Janeiro 2016 e PyeongChang 2018. O Gunning Fog Index, foi aplicado para analisar uma amostra de 30 tweets emitidos por sistema automatizado e 30 tweets emitidos por jornalistas, em cada um dos eventos. Os resultados indicam que a escrita desta IA é simples e concisa, usando apenas uma frase por tweet. As classes gramaticais encontradas foram: substantivos, verbos e adjetivos. No Rio de Janeiro, em 2016, tanto jornalistas (10,54) quanto o bot (10,44) emitiram notas que podem ser lidas sem dificuldade, de acordo com o Gunning Fog Index. O resultado mais significativo foi encontrado na análise de legibilidade das notas emitidas em PyeongChang 2018, onde os jornalistas alcançaram um valor de 14,63 contra 10,8 do OlyBot, o que indica, segundo o modelo utilizado, que se trata de textos de difícil leitura e não de consumo popular. À medida que o número de palavras aumentava, o bot de IA mantinha seu nível de legibilidade em ambos os eventos, enquanto o jornalista aumentava a complexidade de suas notas, tornando-as mais ininteligíveis. As implicações dessas reflexões interessam a tecnólogos, jornalistas e linguistas, bem como à sociedade em geral, pois apresentam o desafio de discernir e avaliar aqueles textos que atualmente podem ser criados por robôs ou por jornalistas, sem comprometer a legibilidade de sua leitura.

Palavras-chave: inteligência artificial, jornalismo automatizado, análise do discurso, redes sociais, bots

Introducción

El periodismo tiene una importante función en la sociedad, pues constituye la manera en que todas las personas pueden estar al día con las últimas novedades, enterándose de manera escrita, oral o auditiva. En palabras de Martini (2000), las noticias “construyen una parte de la realidad social, [...] posibilitan a los individuos el conocimiento al que no pueden acceder de manera directa” (p. 1).

Desde la invención de la imprenta hasta la era de la inteligencia artificial (IA), el periodismo ha transitado por muchos cambios para cumplir con su propósito, y para satisfacer las necesidades y demandas de la sociedad. En el marco de los avances tecnológicos recientes, la IA ha tenido notables progresos en diversas áreas de la sociedad, y el periodismo no ha quedado al margen de este fenómeno. Un ejemplo palpable en este ámbito es la aparición de herramientas automatizadas para la generación de contenidos, llamadas bots. Latar (2018) afirma que el periodismo robótico se inició en los deportes, y luego su práctica se extendió a otros campos, con formas cada vez más sofisticadas.

El periodismo deportivo se ha valido de la inteligencia artificial en los últimos tiempos en eventos de gran cobertura, como los juegos olímpicos de Río de Janeiro, en 2016, y los juegos olímpicos de invierno de 2018, en Pyeongchang. Durante estos eventos, la IA ha sido empleada para mantener las actualizaciones vía Twitter, manejadas por un bot del diario The Washington Post llamado “OlyBot”. Dicha aplicación redactó los tuits acerca de los resultados de las competencias y generó datos sobre la programación de eventos del torneo.

El presente artículo busca encontrar respuestas a la siguiente pregunta: ¿Cuáles son las diferencias discursivas entre las noticias escritas por inteligencia artificial y las noticias escritas por humanos? Se buscará responder a este planteamiento a través del análisis lingüístico de los tuits con la ayuda del índice de niebla de Gunning y el software ATLAS.ti, versión 23, bajo el supuesto de que existen patrones lingüísticos identificables que caracterizan el discurso de los bots automatizados y el de los humanos.

Este trabajo pretende caracterizar el lenguaje que utilizan los bots automatizados, y reflexionar si es posible o no diferenciar cuáles tuits fueron redactados por un algoritmo y cuáles por un humano, a través de la identificación de posibles patrones en las partes del discurso: sustantivos, verbos, adjetivos y adverbios, así como las estructuras lingüísticas identificadas en cada tuit.

Marco referencial

Características del lenguaje generado por inteligencia artificial

La capacidad de un artefacto de reproducir el lenguaje natural se sustenta en el análisis de esta habilidad única del ser humano para comunicarse expresando sentimientos, emociones y pensamientos. Los rasgos que caracterizan el lenguaje humano

propuestos por Hocket (1960) refieren la capacidad de las personas para comunicar, reflexionar sobre el contenido de los mensajes, interpretar palabras y hasta generar significados inéditos a partir de la construcción de infinitas estructuras gramaticales de diferentes niveles de complejidad (Academia Lab, 2023).

El antecedente por excelencia de la comunicación autómatas es el Test de Turing, una prueba diseñada por el matemático Alan Turing para poder evaluar la capacidad de comunicación de una máquina a comparación de un humano. Esta prueba consiste en lo siguiente: “La configuración básica implica tres agentes: un juez y dos participantes. El juez busca saber cuál de los otros dos participantes es una máquina y cuál es un humano” (Kobis y Mossink, 2021, p. 2). En la versión clásica y original del juego, el juez realiza una serie de preguntas en un lapso de cinco minutos; dependiendo de las respuestas de cada participante, el juez dará el dictamen respecto a quién es quién en el juego. En dichas respuestas se patentizan las estructuras gramaticales utilizadas por cada usuario, así como la longitud y complejidad de estas, pudiendo así el juez inferir el origen de cada comunicación.

La producción automática de textos generados por IA también es conocida como “periodismo algorítmico” (Ali y Hassoun, 2019), “periodismo robot” (Firat, 2019), o “periodismo artificial” (Túñez-Lopez et al., 2019). Sea cual sea el término que prefiera utilizarse, debe saberse que se emplean sistemas computacionales para convertir la información y registros en textos. El uso de las tecnologías con IA ha sido más frecuente en fechas recientes, y en muy variadas disciplinas como la medicina, la traducción y las finanzas, entre otras.

El periodismo deportivo, además de poseer una estructura de comunicación identificable y de fácil estandarización, es especialmente vasto en números y estadísticas. Esta información es el insumo principal para los algoritmos que producen las notas periodísticas. Como resultado, no es raro que el periodismo deportivo sea pionero en utilizar la IA en el proceso de la producción de noticias (Gallego, 2021).

Los primeros bots en la web y redes sociales

La década de 1960 marca el inicio a este tipo de tecnologías basadas en IA, ya que es el momento en el que comenzaron a desarrollarse los primeros chats conversacionales, como ELIZA, creada por el Instituto Tecnológico de Massachusetts (MIT), una aplicación que simulaba conversaciones terapéuticas máquina-humano. Si bien sus diálogos y conversaciones son muy limitados, ELIZA constituye la referencia de los chatbots actuales (Lamorte, 2023).

No es sino hasta el comienzo del siglo XXI que estas tecnologías comenzaron a incorporarse en las redes sociales y a tener un avance importante en su aplicación. Es cuando el uso de la mensajería instantánea se vuelve más común y recurrente, lo que dio paso a un avance significativo de los bots conversacionales, utilizados especialmente en el ramo de atención al cliente (Lamorte, 2023). Estos sistemas están configurados para facilitar respuestas automatizadas a cualquier duda de los clientes sobre algún producto o

servicio. El avance de esta tecnología fue más notorio cuando se comenzaron a emplear funciones de voz y no únicamente de texto.

En el año 2011 aparece Siri, una aplicación desarrollada por la empresa estadounidense Apple. Siri utiliza el procesamiento de lenguaje natural para dar respuestas por voz a todo aquello que el usuario le solicite, desde un simple saludo hasta información específica de un tema determinado. Dicha aplicación requiere de la interacción continua del usuario para poder aprender, ampliar su base de datos y refinar sus respuestas. Por lo tanto, mientras mayor sea la cantidad de usuarios utilizando este software, mayor será la inteligencia desarrollada de este bot.

Siri es el sucesor de ALICE, que fue un chat conversacional creado en 1995 por Artificial Linguistic Internet Computer Entity (Marcovich, 2022). La aplicación pretendía interactuar con un humano sin que descubriera que la conversación estaba siendo llevada a cabo con un bot. Es a partir de esta idea que da origen al principio de operación de Siri: la adquisición y el uso del lenguaje natural. Alexa es otro asistente virtual creado en 2014 por la empresa Amazon. El dispositivo realiza las mismas funciones que Siri, Google Assistant y Cortana. Entre las tareas que puede desarrollar están: reconocer la voz de los usuarios, responder a preguntas específicas de diversos temas y fijar recordatorios (Marcovich, 2022). No obstante, esta IA tiene además la capacidad de hacer compras virtuales cuando el usuario se lo ordene.

Automatización de noticias en el periodismo

La inteligencia artificial ha tenido una influencia importante en el periodismo tradicional, especialmente a través de la automatización en la generación de noticias. Esta automatización es lo que provoca que cada vez más notas periodísticas contengan menos intervención humana en su proceso de creación. La actividad periodística está experimentando una transformación debido a la aparición de los diferentes tipos de bots, como los que generan contenido, los verificadores de datos, chatbots y asistentes virtuales, entre otros, que cumplen diversas tareas, cambiando notablemente el trabajo de los profesionistas.

La automatización de notas se evidencia en procesos periodísticos como los de “priorizar, clasificar y filtrar información” (Carlson, 2015, citado por Túnez-López et al., 2019, p. 3). Sin embargo, la IA en el periodismo no se limita a esas funciones, sino que incluso puede involucrarse en tareas como “métricas de análisis de la audiencia, para determinar temas que cubrir y para, en función de la información obtenida o suministrada en bases de datos, escribir historias” (Carlson, 2015, citado por Túnez-López et al., 2019, p. 3).

La automatización implica una actualización tecnológica de relevancia en el ámbito periodístico, pues requiere perder el miedo a utilizarla y aprovecharla al máximo. A decir de Gary Liu, ejecutivo del South China Morning Post, “capturar y recopilar esos datos no nos aporta nada si no sabemos cómo utilizarlo” (citado por Villareal, 2017, párr. 3). Lo que invita a utilizar dichos algoritmos a favor del desarrollo de la profesión, ya

que, como dice Reg Chua, director editorial de datos de Reuters, “lo verdaderamente importante es la fusión entre humanos y máquinas para trabajar en equipo de forma armoniosa” (Villareal, 2017, párr. 7).

Chua argumenta que en la automatización “el enfoque debe ser mostrar datos y tendencias que ayuden a las noticias, ya que esa será la única forma de impulsar la calidad y la eficiencia del medio” (Villareal, 2017, párr. 5). Aunado a esto, otros de los beneficios serían: “la velocidad con la que se crean los nuevos contenidos, el alcance en la creación de contenidos e, incluso, la rentabilidad” (Villareal, 2017, párr. 5). Es la automaticidad periodística lo que permitiría a los profesionales crear noticias personalizadas a favor de su labor y de los intereses del consumidor y, de esta manera, tener un mayor alcance con su trabajo y un mayor consumo por parte del lector.

El pionero en utilizar la IA fue Los Angeles Times (Ufarte Ruiz y Manfredi Sánchez, 2019). En 2014, el periódico creó un software con el objetivo de informar de manera inmediata sobre los últimos eventos sísmicos en Estados Unidos; esta IA fue llamada Quakebot. Este suceso constituye el punto de partida para la industria periodística en la utilización de este tipo de tecnologías en su labor.

En ese mismo año, la agencia estadounidense Associated Press (AP) desarrolló una tecnología para producir crónicas y noticias de beisbol utilizando las estadísticas arrojadas durante los partidos (Bullido, 2019). De esta forma, los bots empiezan a incursionar en la rama del periodismo deportivo. No obstante, la agencia AP no fue la única en explorar esta disciplina, también Reuters y Bloomberg son dos agencias de noticias pioneras en emplear este tipo de tecnologías en sus oficinas, al día de hoy, el 35% de sus noticias son producidas por la IA (Bullido, 2019).

Metodología

Este estudio es de enfoque mixto, el cual se define como “un proceso que recolecta, analiza y vincula datos cuantitativos y cualitativos en un mismo estudio o una serie de investigaciones para responder a un planteamiento” (Ruiz et al., 2013, citados por Enríquez y Argota, 2016, p. 4). La investigación pretende medir e identificar las características discursivas de los tuits emitidos por una IA y por periodistas humanos durante los eventos mundiales de los juegos olímpicos Río de Janeiro 2016 y los juegos de invierno Pyeongchang 2018.

Para llevar a cabo este estudio, se utilizó el método para medición de legibilidad de textos llamado Índice de niebla de Gunning (FOG Index), que fue creado en el año de 1944 (Silva, 2019) y de esta manera medir la inteligibilidad, según los parámetros del índice, de los tuits de diferentes competencias redactados, ya sea por IA o por un periodista en estas dos competencias deportivas.

En la literatura existen muchas herramientas que se utilizan para verificar la facilidad de lectura de los textos. El índice de niebla de Gunning es un modelo clásico

que se sigue utilizando en la lingüística moderna para identificar la dificultad de un texto, evaluando su facilidad de lectura (Ragins, 2012; Sujatna et al., 2021).

No existe una discusión sólida sobre las limitaciones del índice de Gunning en tuits, solo hay evidencia de estudios que han aplicado este modelo en textos breves como ensayos o comentarios de encuestas de satisfacción (Lezner, 2014; Deslongchamps, 2018). El índice de Gunning toma en cuenta la complejidad y la longitud de las palabras de un texto, lo que podría constituir una limitación sobre la aplicación del modelo en textos breves como los tuits por tratarse en realidad de mensajes muy cortos, donde se reduce el margen de análisis para determinar su complejidad de lectura (Pentel, 2015).

Finalmente, se usó el software ATLAS.ti v.23, con el cual se analizó la frecuencia con que las distintas partes del discurso se usaron en cada contexto: sujeto, verbo, adjetivo, adverbio, así como la longitud de estos.

Análisis lingüístico

Índice de niebla de Gunning

El índice de niebla de Gunning es una prueba creada por Robert Gunning, quien, como dice Scott (2023), “observó que la mayoría de los graduados de preparatoria no sabían leer. Gran parte de este problema de lectura era por un problema de escritura. Su opinión era que los periódicos y los documentos empresariales estaban llenos de ‘niebla’ y complejidad innecesaria” (párr. 2).

Este algoritmo de legibilidad de textos se diseñó para el idioma inglés; el índice se calcula con la siguiente fórmula (citado por Silva, 2019, párr. 18):

$$\text{Índice de niebla de Gunning: } 0.4 \times \left(\frac{\text{total de palabras}}{\text{total de oraciones}} \right) + 100 \left(\frac{\text{palabras complejas}}{\text{total de palabras}} \right)$$

En otras palabras, el índice de niebla de Gunning se mide la siguiente forma (Ovares Barquero y Rubí Barquero, 2010):

1. Elegir un texto de más de 100 palabras.
2. Calcular el promedio de la longitud de cada frase (dividir el número de palabras totales por el número de frases en el fragmento).
3. Contar las palabras complejas, aquellas que tienen tres o más sílabas.
4. Agregar la longitud media de las oraciones y el porcentaje de palabras complejas.
5. Multiplicar el resultado por 0.4.

Los resultados del Índice de niebla de Gunning pueden variar entre los valores de 6-17, teniendo decimales como posibles resultados, y cada número entero tiene un significado en específico, como se muestra en la Tabla 1.

Tabla 1. Índice de niebla de Gunning, nivel de lectura por grado de educación norteamericana (Silva, 2019, párr. 15).

Índice de niebla	Nivel de lectura por grado de educación norteamericana	Tipo de publicaciones
17	Graduado universitario	Publicaciones académicas
16	Cuarto año de universidad	Publicaciones médicas, que solo entenderían los graduados.
15	Tercer año de universidad	Publicaciones que no son para consumo popular, difíciles de leer
14	Segundo año de universidad	Publicaciones que no son para consumo popular, difíciles de leer
13	Primer año de universidad	Publicaciones que no son para consumo popular, difíciles de leer
Esta es la línea de peligro, de aquí hacia arriba los textos son muy difíciles de leer		
12	Cuarto año de bachillerato	Periódicos especializados en economía, política
11	Tercer año de bachillerato	Revistas de reportajes más profundos
Todo lo que está debajo de esta línea es fácil de leer.		
10	Segundo año de bachillerato	Publicaciones estilo <i>National Geographic</i>
9	Primer año de bachillerato	Revista como <i>Reader's Digest</i>
8	Octavo grado	Periódicos locales
7	Séptimo grado	La Biblia, libros de Mark Twain
6	Sexto grado	Revistas para público en general, como <i>Hola</i>

Fuente: Silva (2019, párr. 15).

Hallazgos

La IA llamada OlyBot, que opera en la plataforma Twitter (ahora “X”), comienza sus publicaciones en el año 2016 con la cobertura de los Juegos Olímpicos Río de Janeiro, Brasil 2016. En estas publicaciones, OlyBot informó sistemáticamente los resultados de las competencias, como medallas obtenidas, futuros encuentros y posibles cambios de horario de diversos eventos.

En el análisis de los 30 tuits seleccionados para el presente estudio se puede observar que la redacción de esta IA es simple y concisa, pues hace uso de únicamente una oración por tuit. Las partes del discurso más comunes encontradas en sus notas fueron: sustantivo, verbo y adjetivo.

Entendido el Índice de niebla de Gunning como una herramienta para detectar el nivel de dificultad de textos mayores a 100 palabras, en general, los resultados indican que en Río de Janeiro 2016 ambos, periodistas (10.54) y bot (10.44), emitieron notas que se pueden leer sin dificultad (Tabla 2).

El resultado más relevador se puede ver en el análisis del índice de niebla en Pyeongchang 2018, donde las notas de los periodistas alcanzaron un valor de 14.63 contra 10.8 de las generadas por OlyBot (Figura 1), es decir, los periodistas produjeron textos que no son aptos para consumo popular, por su dificultad para leer.

En la transición de Río de Janeiro 2016 a Pyeongchang 2018, OlyBot muestra un claro avance en su redacción, pues ya no se limita únicamente a una oración por tuit, sino

que amplía considerablemente el uso de caracteres (70% más que en Río de Janeiro), y oraciones empleadas, no obstante, esta utilización adicional de palabras no interfirió en la legibilidad de los textos (Tabla 2).

Por su parte, los periodistas utilizaron un 28% de caracteres adicionales en sus notas de Pyeongchang con relación a Río de Janeiro, sin embargo, esta amplitud de palabras también contribuyó a aumentar el índice de legibilidad, lo que propicia que los textos sean más difíciles de leer (Tabla 2).

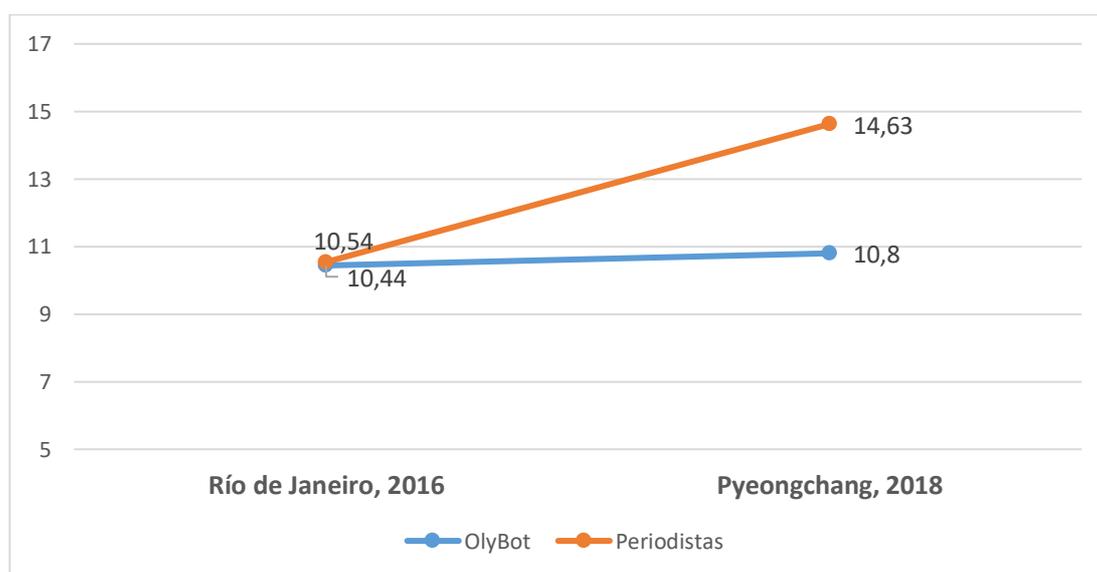
Si bien el número de oraciones y palabras empleadas en los tuits de los periodistas es menor al de las empleadas en los tuits de OlyBot, los tuits de periodistas exhiben una gran diferencia. Esto debido a las partes del discurso identificadas en los tuits, pues de un evento a otro ya no se limitan únicamente a sustantivos, verbos y adjetivos, sino que ahora incluyen adverbios en mayor cantidad de la que hace uso OlyBot. Como resultado, las notas se volvieron más complejas.

Tabla 2. Resultados del índice de niebla de Gunning.

	Río de Janeiro, 2016		Pyeongchang, 2018	
	OlyBot	Periodistas	OlyBot	Periodistas
Legibilidad	10.44	10.54	10.80	14.63
Caracteres	2568	3015	4348	3866
Palabras	403	452	697	659
Oraciones	30	60	49	28
Sílabas	652	746	1132	962

Fuente: Elaboración de los autores.

Figura 1. Nivel de legibilidad de los tuits generados por OlyBot y por periodistas en ambos eventos.



Empleando el software ATLAS.ti v.23, se identificaron las 10 palabras más utilizadas por OlyBot (Tabla 3), de las cuales constaron de dos verbos que ocuparon los primeros dos lugares en la tabla de palabras usadas con mayor frecuencia (“wins” y “beating”) y que a su vez fueron antónimos; además de 6 sustantivos y 2 adjetivos.

Se entiende de esta manera que el algoritmo seguía un patrón definido (propio de los autómatas) para hacer llegar la información, indicando: “quién” o “quiénes” lo hicieron (atleta o equipo), “qué” se hizo (ganar, perder) y, en algunas ocasiones, “cómo” lo hizo (uso de algún adjetivo para describir la forma de hacer las cosas o dar algún detalle respecto al contexto).

Tabla 3. Las 10 palabras más usadas por OlyBot y Periodistas en Río de Janeiro 2016, en orden de frecuencia (30 tuits seleccionados).

OlyBot					Periodistas				
Pos	Palabra	Longitud	Frecuencia	Partes del discurso	Pos	Palabra	Longitud	Frecuencia	Partes del discurso
1	Wins	4	30	Verbo	1	Medal	5	12	Sustantivo
2	Beating	7	28	Verbo	2	#Silver	7	11	Sustantivo
3	Gold	4	21	Sustantivo	3	#Gold	5	8	Sustantivo
4	Men’s	5	17	Adjetivo	4	#Bronze	7	6	Sustantivo
5	#GBR	4	16	Sustantivo	5	You	3	5	Sustantivo
6	#USA	4	13	Sustantivo	6	Is	2	5	Verbo
7	Women’s	7	11	Adjetivo	7	Second	6	4	Adjetivo
8	Bronze	6	9	Sustantivo	8	Win	3	4	Verbo
9	Gymnastics	10	6	Sustantivo	9	Done	4	3	Verbo
10	Cycling	7	5	Sustantivo	10	Amazing	7	3	Adjetivo

Fuente: Elaboración de los autores.

Tabla 4. Las 10 palabras más usadas por OlyBot y Periodistas en Pyeongchang 2018, en orden de frecuencia (30 tuits seleccionados).

OlyBot					Periodistas				
Pos	Palabra	Longitud	Frecuencia	Partes del discurso	Pos	Palabra	Longitud	Frecuencia	Partes del discurso
1	Gold	4	27	Sustantivo	1	Medal	5	17	Sustantivo
2	Takes	5	22	Verbo	2	Pyeongchang2018	15	13	Sustantivo
3	Medal	5	21	Sustantivo	3	Gold	4	13	Sustantivo
4	Event	5	10	Sustantivo	4	Won	4	13	Verbo
5	Silver	6	20	Sustantivo	5	Olympic	7	10	Sustantivo
6	Wins	4	19	Verbo	6	#TeamCanada	11	8	Sustantivo
7	Bronze	6	16	Sustantivo	7	First	5	6	Adjetivo
8	Time	4	11	Sustantivo	8	Medals	6	5	Sustantivo
9	Women’s	7	10	Adjetivo	9	Wins	4	4	Verbo
10	Men’s	4	10	Adjetivo	10	Run	3	4	Verbo

Fuente: Elaboración de los autores.

Dentro de las 10 palabras más utilizadas por los periodistas en Río de Janeiro (Tabla 3), se encuentran tres partes del discurso: 5 sustantivos (ocupando los primeros lugares en cuestión de palabras usadas con mayor frecuencia), 3 verbos y 2 adjetivos. Nuevamente, es indicativo de que la información presentada por los periodistas era breve y directa, enfocándose en quién o quiénes lo hicieron y qué hicieron.

Por su parte, analizando las palabras más utilizadas por OlyBot en Pyeongchang 2018 (Tabla 4), se puede notar que la cantidad de sustantivos (gold, medal, event, silver, bronze, time) sigue siendo ligeramente mayor en comparación con las otras partes del discurso identificadas en los tuits, los cuales representan el 6%, le siguen los verbos, que representan un 2%, y los adjetivos, también con 2%. Cabe destacar que, dentro de todos los tuits analizados, no solamente se identifican estas tres partes del discurso, sino que ahora la IA también hace uso de adverbios, que, si bien son pocos los utilizados, este hecho muestra el nivel de desarrollo que esta tuvo en el transcurso de dos años.

Por su parte, los periodistas en Pyeongchang 2018 usaron un 6% de sustantivos, los cuales ocupan los tres primeros lugares de la tabla, seguido de 3% verbos y 1% adjetivos.

Tabla 5. Comparación de los resultados del conteo de los elementos iniciales de la estructura sintáctica de los tuits.

	Río de Janeiro, 2016		Pyeongchang, 2018	
	OlyBot	Periodistas	OlyBot	Periodistas
Sustantivos	30	18	13	18
Verbos	0	1	0	0
Adjetivos	0	2	16	6
Adverbios	0	0	1	0
Otros	0	9	0	6

Fuente: Elaboración de los autores.

Analizando la estructura sintáctica de OlyBot en el año 2016, podemos darnos cuenta de que todos los tuits de esta IA comenzaron utilizando la misma parte del discurso: el sustantivo, ya que siempre mencionaba al país como el sustantivo de la oración (Tabla 5). Cuando un texto comienza utilizando un sustantivo, implica que la atención del lector se vaya directamente a quién realizó la acción, en lugar del cómo, cuándo, dónde se realizó.

En el caso de OlyBot en el año 2018, cuando la red social ya había presentado una modificación en la cantidad de caracteres para ser utilizados (de 140 a 280 caracteres por tuit), la redacción de la IA se caracterizó por empezar los tuits utilizando un adjetivo. Los adjetivos son aquellas palabras que dan detalles sobre el sustantivo de la oración; en este caso, los adjetivos utilizados fueron referentes a las nacionalidades de los atletas, haciendo énfasis en los países que habían sido los ganadores de la medalla de oro.

Por otra parte, los periodistas, tanto en el 2016 como en el 2018, conservaron el mismo patrón sintáctico a la hora de empezar sus actualizaciones, incluso utilizaron sustantivos con la misma frecuencia para iniciar los tuits.

Contraria a la técnica ampliamente utilizada por los periodistas en las notas periodísticas o columnas, de utilizar el verbo para captar atención en la acción del sujeto, en este estudio, se encontró que los tuits generados por inteligencia artificial, así como los redactados por periodistas, siguieron un patrón predecible, comenzando por el sustantivo en cuestión, haciendo así saber directamente quién hizo qué en las competencias que se llevaron a cabo.

Discusión

La inteligencia artificial está cambiando el oficio periodístico. Los periodistas temen perder su trabajo debido a la automatización de la información que, por otra parte, les aligera la tarea de procesar grandes cantidades de datos (Brousard et al., 2019, citado en Canavilhas, 2022). Para comprender mejor este fenómeno, es necesario analizar los retos, oportunidades e impacto que surgen por la implementación de la IA en el periodismo.

En el periodismo, la transformación digital se refiere a la incorporación de tecnologías en esta disciplina, incluyendo el concepto del periodismo automatizado, es decir, la generación de contenido periodístico de forma automática a través de algoritmos. Aunado a ello, está el tema de la dependencia cada vez más fuerte de plataformas tecnológicas, y cómo afecta la independencia editorial de los medios (Peña-Fernández et al., 2023).

En este panorama, es fundamental considerar el concepto de ética periodística. Algunos trabajos (Ventura-Pociño, 2022, citado en Canavilhas y Giacomelli, 2023) exponen nuevas inquietudes con referencia a los contenidos generados por autómatas, en cuanto a su coherencia con los enfoques editoriales, y que además preserven los valores que rigen el periodismo, como el respeto por la pluralidad y la secrecía de la información de los usuarios, entre otros (Canavilhas y Giacomelli, 2023). En este sentido, la UNESCO desarrolló la norma sobre la ética de la IA: la "Recomendación sobre la ética de la inteligencia artificial", la cual es aplicable en los 194 Estados miembros de la UNESCO a partir de noviembre de 2021 (UNESCO, 2023).

En este trabajo se clasifica el discurso de los bots y de los periodistas en categorías léxicas que se refieren a cada una de las partes del discurso en una oración, a saber: sustantivos, adverbios, adjetivos y preposiciones, entre otros. Clasificación que se basa en sus propiedades, tanto sintácticas como semánticas (Alcaraz Varó y Martínez Linares, 2004).

En 1952, Robert Gunning ofreció las pautas para analizar la facilidad de lectura de los textos, y proporcionó recomendaciones para la escritura clara y efectiva en su libro *The Technique of Clear Writing*. Para Gunning, un texto fácil de leer debe priorizar lo

simple ante lo complejo, utilizar palabras de uso común, evitar palabras innecesarias y escribir para expresar, no para impresionar (Gunning, 1952).

Para una comunicación del tipo notas periodísticas deportivas, el rango ideal del índice de niebla de Gunning es de 6-10. Por encima de este índice, los textos se consideran de una dificultad media para la mayoría de las personas. Las comunicaciones con un índice por encima de 10 corresponden a publicaciones con mayor profundidad en su redacción, o textos especializados en una disciplina en particular (Silva, 2019).

En el caso de la presente investigación, los resultados indican que en Río de Janeiro 2016 ambos, periodistas (10.54) y bots (10.44), emitieron notas que quedaron apenas por encima del rango de textos que se pueden leer sin dificultad, lo cual es aceptable para este tipo de comunicaciones. Sin embargo, en Pyeongchang 2018 se pudo constatar que los periodistas alcanzaron un índice de 14.63 contra 10.8 de OlyBot, lo que indica, según el modelo de Gunning, que se trata de textos que no son para consumo popular, difíciles de leer (Figura 1). Por su parte, OlyBot se mantuvo con un nivel de legibilidad fácil, apto para personas con nivel educativo entre segundo y tercero de bachillerato, según el modelo.

Es importante señalar que la plataforma Twitter (“X”) tuvo un cambio importante a finales del año 2017. Dicho año marcó historia para los usuarios de la red social, pues se dio la oportunidad de incrementar de 140 a 280 los caracteres por tuit (Pymediar, 2023), dando así la oportunidad de abordar más información por tuit del usuario. Debido a que esta modificación en la plataforma se realizó entre los Juegos Olímpicos del 2016 y los Olímpicos de Invierno del 2018, OlyBot y los periodistas seleccionados tuvieron la oportunidad de crear un contenido más específico y detallado.

Sin embargo, ante la posibilidad de escribir con un mayor número de palabras, al parecer los periodistas eligieron utilizar palabras menos comunes, agregando una dificultad innecesaria a los textos, mientras que la inteligencia artificial, como es de esperarse, mantuvo una línea discursiva simple, agregando mayor información a los comunicados.

Este resultado es comprensible, dado que la IA generadora de contenidos está diseñada para crear textos basados en un gran cuerpo de recursos escritos para predecir patrones de escritura (Franganillo, 2023). Por esta característica, la IA es capaz de crear contenido nuevo con textos altamente predecibles, de tal manera que, si a un agente de IAG se le instruye para crear textos con características definidas, como estructura, longitud del mensaje, contexto y hasta tono del comunicado, el bot se apegará a estas características, y su estilo no variará con el aumento de caracteres permitidos.

Conclusiones

Se habla mucho de la inmensa capacidad de la IA de aprender y desarrollar las capacidades para lo que fue creado, a partir de la interacción con los seres humanos. Con los datos recopilados para este análisis puede notarse la evolución de OlyBot tuvo a lo largo de dos años, pues no solamente mejoró la longitud de las oraciones con las que presentaba información para dar más datos, sino que empleó estructuras lingüísticas variadas, así como una nueva parte del discurso que no fue empleada en el año 2016 pero sí en el 2018: el adverbio.

En todos los casos, los tuits tuvieron la tendencia de comenzar con un sustantivo, mostrándose directos en informar quién había obtenido un logro. Sin embargo, en el año 2018, OlyBot muestra una evolución, la cual es incorporar los adjetivos como la primera palabra en ser utilizada para dar información más detallada acerca del sujeto que había ganado algo.

El resultado más revelador del presente estudio radica en la característica principal de los sistemas automatizados: la capacidad para mantener la sistematicidad de su labor. Es notorio que el bot pudo conservar su nivel de legibilidad o facilidad de lectura a pesar de incrementar considerablemente el número de caracteres por redacción de un evento a otro. Por su parte, el periodista utilizó también una mayor cantidad de caracteres por tuit al paso de dos años, sin embargo, esta alteración incidió drásticamente en el nivel de legibilidad de sus notas, haciéndolas más complejas y difíciles de leer.

Las implicaciones de estas reflexiones son aplicables a los tecnólogos y lingüistas, así como para la sociedad en general, pues se presenta el reto de discernir aquellos textos que son creados por automatización o por periodistas, para evitar ser sorprendidos por noticias falsas o en diversas situaciones que provoquen la desinformación.

Referencias

- AcademiaLab. (2025). *Charles F. Hockett*. Enciclopedia. <https://academia-lab.com/enciclopedia/charles-f-hockett/>
- Alcaraz Varó, E., & Martínez Linares, M. A. (2004). *Diccionario de lingüística moderna*. Editorial Ariel.
- Ali, W., & Hassoun, M. (2019). Artificial Intelligence and Automated Journalism: Contemporary Challenges and New Opportunities. *International Journal of Media, Journalism and Mass Communications*, 5(1), 40–49. <https://doi.org/10.20431/2454-9479.0501004>
- Bullido, A. (2019). #54 *El auge de los bots en periodismo: ¿oportunidad o peligro?* Escuela de Periodismo. <https://open.spotify.com/episode/3WHna8IiNHuL7nMkHIZWV>

- Canavilhas, J., & Giacomelli, F. (2023). Artificial Intelligence in sports journalism: a study in Brazil and Portugal. *Revista de Comunicacion*, 22(1), 53–69. <https://doi.org/10.26441/RC22.1-2023-3005>
- Canavilhas, J. (2022). Artificial Intelligence and Journalism: Current Situation and Expectations in the Portuguese Sports Media. *Journalism and Media*, 3(3), 510–520. <https://doi.org/10.3390/journalmedia3030035>
- Deslongchamps, A. (2018). Readability and the Bank of Canada. Bank of Canada. <https://www.banqueducanada.ca/2018/06/note-analytique-personnel-2018-20/>
- Enríquez Salas, P., & Argota Pérez, G. (2016). Descripción interpretativa para la elaboración del perfil de tesis de investigación científica con enfoque cualimétrico (mixto). *Campus*, 21(22), 151–164. <https://doi.org/10.24265/campus.2016.v21n22.02>
- Firat, F. (2019). Robot Journalism. *The International Encyclopedia of Journalism Studies*, October, 1–5. <https://doi.org/10.1002/9781118841570.iejs0243>
- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methados Revista De Ciencias Sociales*, 11(2), m231102a10. <https://doi.org/10.17502/mrcs.v11i2.710>
- Gunning, R. (1968). The technique of clear writing. In 1952. <https://dokumen.pub/qdownload/the-technique-of-clear-writing-1nbsped-9780070252066.html>
- Hockett. (1982). Hockett, Charles F. (1960) The Origin of Speech. *Scientific American*, 1960, 4–12. <https://web.stanford.edu/class/linguist197a/hockett60sciam.pdf>
- Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114(106553), 1–13. <https://doi.org/https://doi.org/10.1016/j.chb.2020.106553>
- Lamorte, J. M. (2023). *Bots, las herramientas omnipresentes en la era digital actual*. Next Work 360. <https://www.itmastersmag.com/inteligencia-artificial/bots-las-herramientas-omnipresentes-en-la-era-digital-actual/>
- Latar, N. (2018). *Robot Journalism: Can Human Journalism Survive?* World Scientific Publishing. https://www.academia.edu/77507591/Robot_Journalism_Can_Human_Journalism_Survive
- Lenzner, T. (2014). Are Readability Formulas Valid Tools for Assessing Survey Question Difficulty? *Sociological Methods and Research*, 43(4), 677–698. <https://doi.org/10.1177/0049124113513436>

- López Gallego, J. (2021). *Los bots y el nuevo periodismo automatizado*. Aula En Red Acción. <https://blogotomia.com/aulaenred-accion/los-bots-el-nuevo-periodismo-automatizado/>
- Marcovich, D. (2022). *La Historia de los Chatbots*. AutomatiChat. <https://automaticchat.com/blog/la-historia-de-los-chatbots>
- Martini, S. (2000). *Periodismo, noticia y noticiabilidad*. Editorial Norma.
- Ovares Barquero, F. (2010). Variación del índice de niebla usando un corpus obtenido a partir de los libros digitalizados por google. *Uniciencia*, 24, 133–141.
- Pentel, A. (2015). A Comparison of Different Feature Sets for Age-Based Classification of Short Texts. 6th International Conference on Information, Intelligence, Systems and Applications (IISA). <https://ieeexplore.ieee.org/document/7388069>
- Peña-Fernández, S., Meso-Ayerdi, K., Larrondo-Ureta, A., & Díaz-Noci, J. (2023). Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media. *Profesional de La Información*, 32(2), 1–16. <https://doi.org/10.3145/epi.2023.mar.27>
- Pymediar. (2023). *Twitter y su historia: cómo se convirtió en X*. Pymediar.Com/Blog. <https://www.pymediar.com/twitter-y-su-historia-como-se-convirtio-en-x/#:~:text=>
- Ragins, B. (2012). Editor's comments : Reflections on the Craft. *Academy of Management Review*, 37(4), 493–501.
- Scott, B. (2025). *The Gunning Fog Index (or FOG) Readability Formula*. Readability Formulas. <https://readabilityformulas.com/the-gunnings-fog-index-or-fog-readability-formula/>
- Silva, A. (2018). *Mejora tu SEO: Qué es y cómo se calcula el índice de niebla de Gunning-Fog*. PWS. <https://www.posicionamiento-web-salamanca.com/blog/seo/mejora-seo-calcula-indice-niebla/>
- Sujatna, E. T. S., Heriyanto, H., & Andri, S. (2021). Lexical density and variation in Indonesian folklores in English student textbooks: an SFL study. *Leksika: Jurnal Bahasa, Sastra Dan Pengajarannya*, 15(2), 62. <https://doi.org/10.30595/lks.v15i2.11102>
- Túñez-López, M., Toural-Bran, C., & Valdiviezo-Abad, C. (2019). Automatización, bots y algoritmos en la redacción de noticias. Impacto y calidad del periodismo artificial. *RLCS, Revista Latina de Comunicación Social*, 74, 1411–1433. <https://doi.org/10.4185/RLCS>
- Ufarte Ruiz, M. J., & Manfredi Sánchez, J. L. (2019). Algorithms and bots applied to journalism. The case of Narrativa Inteligencia Artificial: Structure, production

and informative quality. *Doxa Comunicacion*, 2019(29), 213–233.
<https://doi.org/10.31921/doxacom.n29a11>

UNESCO. (2023). *UNESCO's Recommendation on the Ethics of Artificial Intelligence*.
Unesco. <https://unesdoc.unesco.org/ark:/48223/pf0000385082.page=4>

Villarreal, M. (2017). *El impacto de la IA y la automatización sobre las noticias*.
Medialab. <https://medialab.press/impacto-la-ia-la-automatizacion-las-noticias/>