

Artículo de Investigación

Mediación algorítmica: sesgos en la búsqueda de información en estudiantes universitarios

Algorithmic mediation: bias in the information search by university students

Martha Vidal-Sepúlveda: Universidad Central de Chile, Chile.

martha.vidal@uccentral.cl

Cristian Olivares-Rodríguez¹: Universidad Alberto Hurtado, Chile.

colivares@uahurtado.cl

Fecha de Recepción: 26/05/2024

Fecha de Aceptación: 16/07/2024

Fecha de Publicación: 19/07/2024

Cómo citar el artículo (APA 7^a):

Vidal-Sepúlveda, M. y Olivares-Rodríguez, C. (2024). Mediación algorítmica: sesgos en la búsqueda de información en estudiantes universitarios [Algorithmic mediation: bias in the information search by university students]. *European Public & Social Innovation Review*, 9, 1-18. <https://doi.org/10.31637/epsir-2024-397>

Resumen:

Introducción: La búsqueda de información en internet está sujeta a las reglas de la web, requiriendo de la mediación de un algoritmo de búsqueda. En este estudio proponemos caracterizar la interacción del pensamiento crítico y la saturación de información sobre las decisiones de estudiantes de primer ingreso en procesos de búsqueda de información, permitiendo reconocer los sesgos cognitivos presentes en dichas decisiones. **Metodología:** Se realizó un estudio experimental con estudiantes de primer año, quienes resolvieron tres tareas de búsqueda de información. La plataforma GoNSA2, recopiló automáticamente las trazas del proceso de búsqueda para cada tarea. **Resultados:** Los principales hallazgos refieren a la operación de los sesgos cognitivos de anclaje, *ranking* y disponibilidad, los que dependen de la operación del pensamiento crítico, la saturación de información y del contexto de las tareas. **Discusión:** Se evidencia que la operación de los sesgos depende de la interacción conjunta del pensamiento crítico, la saturación de información y del contexto de las tareas, por lo que el contexto emerge como un determinante de las decisiones de los estudiantes. **Conclusiones:** Estos hallazgos nos permiten sustentar la importancia de modelos de búsqueda que habiliten nuevas estrategias pedagógicas orientadas a fomentar la competencia crítica de búsqueda de información.

¹ **Autor Correspondiente:** Cristian Olivares-Rodríguez. Universidad Alberto Hurtado (Chile).

Palabras clave: búsqueda de información; mediación algorítmica; sesgos; pensamiento crítico; saturación; anclaje; disponibilidad; estudiantes universitarios.

Abstract:

Introduction: Internet information search is governed by web rules, requiring the mediation of a search algorithm. In this study, we aim to characterize the interaction between critical thinking and information saturation on the decisions of first-year students during information search processes, enabling the recognition of cognitive biases present in these decisions. **Methodology:** An experimental study was conducted with first-year students who completed three information search tasks. The GoNSA2 platform automatically recorded the search process traces for each task. **Results:** The main findings pertain to the operation of cognitive biases such as anchoring, ranking, and availability, which depend on the operation of critical thinking, information saturation, and the task context. **Discussion:** It is evident that the operation of biases depends on the combined interaction of critical thinking, information saturation, and the task context, making context a determinant of students' decisions. **Conclusions:** These findings support the importance of search models that enable new pedagogical strategies aimed at fostering critical information search competence.

Keywords: information search; algorithmic mediation; biases; critical thinking; saturation; anchor; availability; university students.

1. Introducción

Para solucionar una necesidad de información accedemos a internet a través de nuestro motor de búsqueda favorito, desde cualquier lugar con conexión a internet y desde diferentes dispositivos tecnológicos. La búsqueda de información en internet está sujeta a las reglas de la web, requiriendo de la mediación de un algoritmo de búsqueda. En internet existen diferentes motores de búsqueda y cada uno de ellos cuenta con una mediación algorítmica basada en sus propios criterios de indexación y recuperación de información. De esta forma, no ha sido posible determinar si, ante una consulta idéntica, es probable obtener la misma fuente de información o el mismo orden de relevancia en diferentes motores de búsqueda (Yagci *et al.*, 2022). Actualmente, estos algoritmos son capaces de inducir modelos generalizados de comportamientos o fenómenos, a partir de grandes cantidades de datos (Baeza-Yates, 2018). Existe una corriente de investigación preocupada por las implicancias de los modelos computacionales inferidos desde datos masivos sobre la vida cotidiana de las personas, debido a su impacto sobre el comportamiento de los usuarios (Chander, 2016). Lo anterior, ha incrementado su tratamiento en la literatura, traspasando las fronteras de la informática para ser un objeto de interés de las ciencias sociales, principalmente en lo relativo al uso, manipulación y privacidad de la información (Burrell, 2016, p.2). Dichas decisiones algorítmicas han sido definidas como opacas (Pasquale, 2015), debido a que no es posible conocer con exactitud las decisiones del modelo inferido desde los distintos algoritmos, por lo que las personas ignoramos las reglas algorítmicas aplicadas sobre los datos y, en particular, las que seleccionan nuestros resultados en el motor de búsqueda. En consecuencia, la mediación algorítmica de estos dispositivos tecnológicos acota los resultados de manera transversal a todas las personas, limitando el acceso a la información disponible en función de sus decisiones históricas, la geolocalización y el idioma entre otros criterios (Olivares-Rodríguez *et al.*, 2021; Vidal-Sepúlveda *et al.*, 2021).

En este ambiente digital, el acceso a la información depende de la selección del algoritmo de búsqueda, devenido en curador mecánico de contenidos (Bhaskar, 2018). Esto supone la oportunidad de acceder a fuentes de conocimiento de difícil acceso antes de la digitalización,

asumiendo que el vasto uso de la tecnología como medio de comunicación faculta a las personas como expertas en la búsqueda de información. Sin embargo, tienden a confiar en la tecnología, delegando sus decisiones en las recomendaciones del algoritmo de su buscador favorito, por ejemplo, aceptando como válido el resultado que más se repite (Ghenai *et al.*, 2020).

Durante la pandemia, la educación avanzó forzosamente a modelos de enseñanza virtual e híbrida, lo que aceleró la irrupción de las TIC en las instituciones educativas, dejando manifiesta las brechas digitales de docentes y estudiantes (Suárez-Guerrero y Lloret-Catalá, 2022). En general, para el estudiantado, comunicarse y aprender de manera ubicua es una actividad integrada en la vida cotidiana (Burbules, 2012). En este sentido, y tal como lo plantea Pereira *et al.* (2019), las personas participan activamente del entramado mediático global desde edades cada vez más tempranas. No obstante, en su estudio evidencia que los medios aún no están integrados en el aula. Por lo tanto, quedan expuestas a la sobrecarga de información, la desinformación, las noticias falsas, entre otros desórdenes informativos. Esto supone la necesidad urgente de formarlos para la selección cuidadosa de la información. Sin embargo, desde las instituciones educativas no existen planes de formación de alfabetización crítica que les permitan seleccionar fuentes fiables de información ni cuestionar el contenido de la información a la que acceden en la web. Por ejemplo, Sales *et al.* (2020) evidencia que el profesorado universitario español manifiesta preocupación respecto de las competencias críticas de sus estudiantes y cuestionan sus propias capacidades para formarlos en dichas competencias. Se ha establecido que no hay avances suficientes en la formación de docentes que incorporen la formación de una perspectiva crítica en la práctica pedagógica (Mesquita-Romero *et al.*, 2022) que permita a las personas desenvolverse como ciudadanos autónomos capaces de realizar una comprensión crítica de los medios y la información disponible (Buckingham, 2008). En general, la alfabetización digital y mediática de la juventud queda a su propia agencia, aprendiendo, principalmente, a partir del ensayo y error, imitación e inspiración y la búsqueda de información (Pereira *et al.*, 2019). En este estudio proponemos caracterizar la interacción del pensamiento crítico y la saturación de información sobre las decisiones de estudiantes de primer ingreso en procesos de búsqueda de información, permitiendo reconocer los sesgos cognitivos presentes en dichas decisiones.

En consecuencia, nos interesa recopilar evidencias de los obstáculos que enfrentan estudiantes universitarios durante la búsqueda de información para resolver tareas académicas. Nuestro objetivo es establecer la relación causal que tiene la interacción entre la saturación de información en los resultados de búsqueda y la motivación al pensamiento crítico, sobre la operación de los sesgos cognitivos durante la búsqueda de información en internet. Por ello, analizamos su comportamiento de búsqueda caracterizando los posibles sesgos y los efectos de la saturación de información, considerando la operación de la motivación de las personas a pensar críticamente. En ese sentido, esta investigación busca responder las siguientes interrogantes:

- ¿Cómo operan los sesgos cognitivos durante la búsqueda de información para resolver una tarea académica?
- ¿La motivación de las personas a pensar críticamente influye significativamente sobre la operación de sesgos cognitivos?
- ¿Las personas que reciben una lista de resultados saturada manifiestan un comportamiento de búsqueda más sesgado?

2. Metodología

Esta investigación se enmarca en un estudio experimental (Campbell y Stanley, 1995) para evaluar la influencia de la interacción entre la saturación de resultados y la motivación al pensamiento crítico sobre los sesgos cognitivos en la búsqueda de información. Se conformó un grupo experimental (en adelante, GE) que recibió una lista de resultados saturada (20 ítems por página), mientras que el grupo de control (en adelante, GC) recibió una lista de resultados tradicional (10 ítems por página). La asignación de participantes a cada grupo se realizó de manera aleatoria siguiendo un proceso estratificado aleatorio en función de las variables de 1) motivación al pensamiento crítico y 2) área disciplinar, lo que permite distribuir la muestra de manera homogénea en los grupos e igualar otras posibles variables extrañas desconocidas por los investigadores (Pick y López, 1998). Ambos grupos realizaron una prueba *ad-hoc* para medir la motivación al pensamiento crítico. El diseño experimental considera a la personalización algorítmica y el área disciplinar como variables extrañas por lo que fueron controladas durante el estudio. La personalización algorítmica se controló mediante un punto único de conexión a través de la API de Bing y, el área disciplinar, considerando a participantes de dos áreas disciplinares.

2.1. Participantes

Se utilizó una muestra no probabilística por conveniencia (Flick, 2007). Los criterios de inclusión de la muestra son: 1) Estudiantes universitarios de primer ingreso, ya que tienen un avance similar en la malla curricular de su carrera y 2) Aceptación voluntaria mediante un consentimiento informado en el que autorizan el tratamiento de sus datos de manera confidencial, anónima y agregada con fines estrictamente académicos. La muestra corresponde a 64 estudiantes universitarios de las carreras de periodismo e ingeniería en informática, con un 51,6% de hombres y 48,4 % de mujeres. La aleatorización de la muestra en grupo de control y experimental (en adelante, GC y GE) se realiza en forma automática a través de cuotas en la plataforma GoNSA2 (Olivares-Rodríguez et al., 2017), la que interactúa con el motor de búsqueda Bing de Microsoft, debido a las libertades legales para registrar de manera local las acciones realizadas por las personas participantes durante la búsqueda de información. Este experimento contó con la aprobación del Comité de Ética y Bioética de la Universidad Austral de Chile.

2.2. Variables independientes

Motivación al Pensamiento crítico: Se utilizó la Escala Motivacional de Pensamiento Crítico, en adelante EMPC (Valenzuela y Nieto, 2008). La EMPC se compone de 20 ítems tipo Likert con puntuaciones de 1 a 6, en la cual las personas expresan su grado de acuerdo o desacuerdo de una serie de afirmaciones respecto de las expectativas que tienen de pensar críticamente y, de afirmaciones valorativas respecto la importancia y utilidad percibida hacia el pensamiento crítico, el costo que asumirán por hacerlo y su interés personal en pensar de forma rigurosa. Así, los ítems de la EMPC son: Expectativa, Valor, Utilidad, Interés y Costo.

Saturación de Información: Corresponde a la sobrecarga de información que reduce la capacidad de tomar decisiones efectivas por parte de las personas (Faulkner y Trotter 2017; Roetzel, 2019). En particular, durante la búsqueda de información esta sobrecarga se evidencia en la cantidad de resultados por página (Kelly y Azzopardi, 2015).

2.3. Variable dependiente

Operación de sesgos cognitivos: Para la identificación de la operación de los sesgos cognitivos se propone observar su operación en los dos puntos de toma de decisiones de las personas participantes: 1) las consultas emitidas y 2) la selección de documentos. La caracterización de cada uno de los sesgos se detalla en el apartado 2.6.

2.4. Diseño de las tareas de búsqueda

Para seleccionar las tareas de búsqueda se elaboró un conjunto de actividades siguiendo los lineamientos de la taxonomía bidimensional de Anderson y Krathwohl (Urgo et al., 2019). Las tareas fueron sometidas a la evaluación de tres personas expertas en: Educación y Tecnología, Neurodidáctica y Lingüística. En el experimento se utilizaron solo las tareas que alcanzaron acuerdo en todas las dimensiones analizadas (Tabla 1).

Tabla 1.

Tareas de búsqueda de información

| Contexto | Consigna de la tarea | Tipo de conocimiento |
|---------------------|---|----------------------|
| Fallecidos COVID 19 | Mencione la cantidad de personas fallecidas por el COVID 19 en Chile durante el año 2020. | Factual |
| Elecciones 2021 | Resuma las razones del cambio de fecha de las elecciones de abril del 2021. | Exploratorio |
| Noticias Falsas | Resuma las características de las noticias falsas. | Exploratorio |

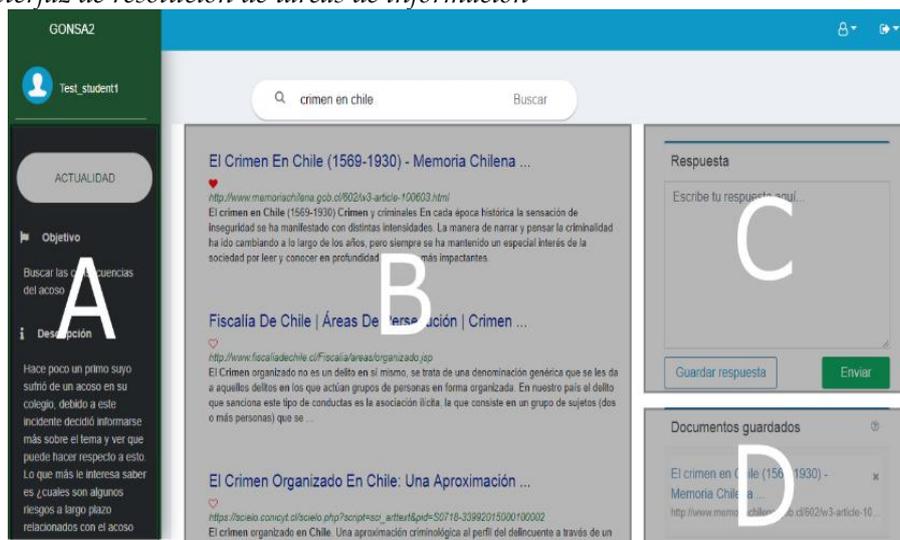
Fuente: Elaboración propia (2024).

2.5. Recolección de datos

La comunicación con las personas participantes se realizó de forma síncrona-*online*. En la primera sesión se aplicó la escala EMPC. Para determinar el nivel de motivación al pensamiento crítico se consideraron las medias de las respuestas en cada dimensión del EMPC por cada estudiante. En la segunda sesión se presentaron las funcionalidades técnicas de la plataforma y se les invitó a resolver las tareas utilizando el buscador de la plataforma (B) en el orden de aparición siguiendo las instrucciones de cada tarea (A) y guardando las páginas webs que consideran relevantes en su biblioteca personal (D). Finalmente, debían enviar su respuesta (C) a la plataforma (Figura 1). Se dispuso de 45 minutos para resolver las tareas.

Figura 1.

GoNSA2: Interfaz de resolución de tareas de información



Fuente: Elaboración propia (2024).

2.6. Caracterización de los sesgos cognitivos

Según Kahneman (2012), el sesgo de anclaje corresponde al uso como punto de referencia del número, valor o información inicial que recibimos. Para determinar el sesgo de anclaje del estudiantado en las tareas de búsqueda analizamos la coincidencia de palabras en las relaciones previamente establecidas en el diseño de análisis. Consideramos que en las personas opera el sesgo de anclaje si existen coincidencias de palabras o términos o, por el contrario, utilizan palabras o términos relacionados. De esta forma, el sesgo de anclaje es la proporción entre el número de palabras contenidas en cada relación y el número total de palabras. Para mejorar la precisión del análisis se eliminaron las tildes y *stopwords* que no pertenecen al léxico nocional (conectores, pronombres, artículos, etcétera).

Durante la sesión de búsqueda, las personas tenían la posibilidad de seleccionar resultados que, a su juicio, fueran relevantes para alcanzar el objetivo de cada tarea, almacenándose en su biblioteca personal. Para caracterizar el sesgo de anclaje fueron considerados sólo los *snippets* de los resultados seleccionados, ya que son la porción de información que influye directamente en la decisión temprana de las personas tanto para explorar como para seleccionar información.

El sesgo de *ranking* corresponde a la creencia extendida entre las personas que los primeros resultados provistos por el buscador tienen mayor relación con la consulta emitida y que se entiende como un comportamiento general de los usuarios (Noble, 2018). En otras investigaciones se ha establecido que las personas seleccionan los resultados que se ubican en las primeras posiciones (Schultheiß et al., 2018; Su et al., 2018) y son influenciadas por el orden de los resultados propuesto por el motor de búsqueda (Novin y Meyers, 2017). Según esto, el sesgo operaría en el estudiantado al seleccionar los primeros resultados, basados en la confianza en la priorización de resultados propuesta por el algoritmo si el título y/o resumen contienen alguna de las palabras de su consulta.

De acuerdo con la literatura, las personas aceptan como respuesta válida la que más se repite entre los resultados de búsqueda (Ghenai et al., 2020). De esta forma, la tendencia es confiar

en una lista de resultados con información homogénea y desconfiar en un listado de información diverso. De esta manera, el sesgo opera si se observa que los participantes reducen el número de consultas ante resultados de búsqueda homogéneos o viceversa.

2.7. Análisis estadístico

Se realizó un análisis descriptivo, así como un análisis de normalidad de las distribuciones de las variables EMPC a través de los estadísticos de Kolmogorov-Smirnov ($p < 0.001$), Shapiro-Wilk ($p < 0.01$) y D'Agostino's K-squared ($p < 0.01$). Mientras que los datos presentan normalidad tanto para el anclaje entre consigna-consulta, consulta-título y consulta-*snippet* para los test Kolmogorov-Smirnov ($p < 0.001$), Shapiro-Wilk ($p < 0.01$) y D'Agostino's K-squared ($p < 0.01$). Dado que las variables cumplen el supuesto de normalidad se decidió aplicar pruebas paramétricas (Hernández-Sampieri, et al., 2018). Por ello, se empleó un análisis a partir del Test U de Mann-Whitney-Wilcoxon para analizar los valores medios de cada tipo de anclaje en función de la interacción entre saturación (GC, GE) y los niveles de motivación al pensamiento crítico de cada subdimensión (Alto, Bajo). La significancia estadística se fijó a un valor $p < 0,05$. El análisis estadístico de los datos fue realizado con las librerías Statsmodels y Scipy de Python.

3. Resultados

3.1. Sesgo de anclaje

El sesgo de anclaje se analizó en dos etapas del proceso de búsqueda, tanto durante la elaboración de las consultas que permiten acceder a la información como en la selección de las fuentes de información entregadas por el motor de búsqueda.

3.1.1. Elaboración y/o reformulación de las consultas:

En esta etapa observamos la operación del sesgo de anclaje en la relación “consigna - consultas” en cada tarea (Tabla 2). En general, se observa que el sesgo de anclaje es mayor en participantes con mayores resultados en la escala EMPC.

Tabla 2.

Sesgo de anclaje entre la consigna - consultas

| Dimensión EMPC | | Covid-19 | | Elecciones 2021 | | Noticias falsas | |
|----------------|------|----------|------|-----------------|------|-----------------|------|
| | | GE | GC | GE | GC | GE | GC |
| Expectativa | Alto | 0,47 | 0,59 | 0,92 | 0,91 | 0,44 | 0,58 |
| | Bajo | 0,41 | 0,51 | 0,74 | 0,74 | 0,42 | 0,51 |
| Importancia | Alto | 0,43 | 0,54 | 0,76 | 0,75 | 0,53 | 0,45 |
| | Bajo | 0,39 | 0,40 | 0,81 | 0,77 | 0,27 | 0,78 |
| Utilidad | Alto | 0,42 | 0,56 | 0,81 | 0,76 | 0,38 | 0,56 |
| | Bajo | 0,44 | 0,39 | 0,62 | 0,77 | 0,58 | 0,37 |
| Interés | Alto | 0,51 | 0,47 | 0,78 | 0,69 | 0,46 | 0,58 |
| | Bajo | 0,36 | 0,59 | 0,79 | 0,83 | 0,39 | 0,42 |
| Costo | Alto | 0,50 | 0,51 | 0,68 | 0,72 | 0,33 | 0,63 |
| | Bajo | 0,41 | 0,53 | 0,8 | 0,78 | 0,44 | 0,46 |

Fuente: Elaboración propia (2024).

Al examinar por tareas, en relación con las “Elecciones 2021” se observa una mayor concentración del anclaje, superando la media de anclaje de las otras tareas (COVID-19 y Noticias Falsas). Por lo tanto, se evidencia una mayor tendencia al anclaje para todo el estudiantado, independientemente del nivel de EMPC reportado. Según esto, el conocimiento previo de la temática de la tarea podría operar en el sesgo de anclaje en tópicos menos conocidos por los participantes, lo que los llevaría a confiar en la información presentada en la consigna. Por otra parte, en el caso de las tareas “COVID-19” y “Noticias falsas”, la capacidad de autorregulación de las personas con mayor EMPC, podría influir sobre la orientación al logro, lo que refuerza la idea de que el EMPC opera cuando el nivel de conocimiento de la tarea por parte de los participantes es menor. Al determinar la influencia de las variables saturación y las sub-escalas de EMPC sobre el anclaje, se observaron diferencias significativas entre los grupos ($p < 0.001$). De la misma manera, el sesgo de anclaje presenta diferencias significativas respecto del contexto de la tarea, así como de las variables de EMPC y saturación. El nivel de anclaje del estudiantado respecto de la consigna de la tarea es significativamente diferente cuando se encuentra saturado de información y en función del tópico de la tarea. Es decir, cuando se enfrentaron la tarea de “Elecciones 2021”, el estudiantado tuvo una fuerte tendencia a anclarse a los términos propuestos en la consigna de la tarea, tales como “elecciones”, “abril”, “2021”, etc.; mientras que en la tarea “COVID-19” tendieron a utilizar mayor diversidad de palabras tales como “muertes”, “defunciones”, entre otras. Lo que evidencia la influencia del tópico de la tarea sobre la selección de términos para las consultas.

3.1.1. Selección de las fuentes de información

Para estudiar la ocurrencia del sesgo de anclaje en la selección de fuentes de información se calcula el número de palabras contenidas en la relación “consulta - título” y en “consulta - snippet” de las fuentes guardadas en la biblioteca del estudiante. En la selección de las fuentes de información respecto del título (Tabla 3) se observan diferencias significativas en la operación del sesgo de anclaje en las distintas tareas. Se observa que el anclaje con el título es menor para la tarea “COVID-19” con un máximo de 0,42. Sin embargo, en las siguientes tareas el anclaje aumenta, con un máximo de 0,7 en la tarea “Elecciones 2021” y un 0,77 en la tarea “Noticias Falsas”. En la tarea “COVID-19” el anclaje con el título es menor en el grupo de control para los estudiantes con alto EMPC. Por el contrario, en las tareas “Elecciones 2021” y “Noticias Falsas” aumenta el anclaje con el título alcanzando la mediana más alta en la tarea “Elecciones 2021” en las personas con alto EMPC. Al determinar la influencia de las variables saturación y las sub-escalas de EMPC sobre el anclaje, no se observaron diferencias significativas entre los grupos ($p > 0.05$).

Tabla 3.
Sesgo de anclaje entre la consulta – título de la biblioteca

| Dimensión EMPC | | Covid-19 | | Elecciones 2021 | | Noticias falsas | |
|----------------|------|----------|------|-----------------|------|-----------------|------|
| | | GE | GC | GE | GC | GE | GC |
| Expectativa | Alto | 0,42 | 0,37 | 0,35 | 0,63 | 0,44 | 0,44 |
| | Bajo | 0,27 | 0,29 | 0,40 | 0,58 | 0,55 | 0,59 |
| Importancia | Alto | 0,28 | 0,29 | 0,35 | 0,59 | 0,54 | 0,55 |
| | Bajo | 0,3 | 0,33 | 0,56 | 0,59 | 0,50 | 0,67 |
| Utilidad | Alto | 0,28 | 0,28 | 0,38 | 0,54 | 0,50 | 0,51 |
| | Bajo | 0,32 | 0,36 | 0,43 | 0,70 | 0,67 | 0,72 |
| Interés | Alto | 0,33 | 0,26 | 0,39 | 0,61 | 0,50 | 0,63 |
| | Bajo | 0,23 | 0,35 | 0,39 | 0,56 | 0,57 | 0,49 |
| Costo | Alto | 0,42 | 0,22 | 0,33 | 0,56 | 0,55 | 0,77 |
| | Bajo | 0,24 | 0,32 | 0,41 | 0,60 | 0,53 | 0,50 |

Fuente: Elaboración propia (2024).

Entre el GE y el GC (Tabla 4) se observan diferencias de distribución en todas las tareas y entre estudiantes con alto y bajo EMPC. Según esto, el estudiantado se ve mayormente influenciado en la selección de las fuentes de información por el título del resultado que por sobre el contenido del *snippet*. Al determinar la influencia de las variables saturación y las sub-escalas de EMPC sobre el anclaje, no se observaron diferencias significativas entre los grupos ($p > 0.05$), salvo la subdimensión de Utilidad ($p < 0.01$).

Tabla 4.
Sesgo de anclaje entre la consulta – snippet de la biblioteca

| Dimensión EMPC | | Covid-19 | | Elecciones 2021 | | Noticias falsas | |
|----------------|------|----------|------|-----------------|------|-----------------|------|
| | | GE | GC | GE | GC | GE | GC |
| Expectativa | Alto | 0,45 | 0,46 | 0,55 | 0,78 | 0,20 | 0,00 |
| | Bajo | 0,40 | 0,48 | 0,54 | 0,58 | 0,35 | 0,36 |
| Importancia | Alto | 0,39 | 0,46 | 0,5 | 0,66 | 0,32 | 0,31 |
| | Bajo | 0,40 | 0,51 | 0,73 | 0,49 | 0,29 | 0,42 |
| Utilidad | Alto | 0,43 | 0,48 | 0,56 | 0,55 | 0,31 | 0,15 |
| | Bajo | 0,28 | 0,46 | 0,43 | 0,71 | 0,33 | 0,75 |
| Interés | Alto | 0,48 | 0,40 | 0,57 | 0,63 | 0,37 | 0,34 |
| | Bajo | 0,32 | 0,56 | 0,50 | 0,57 | 0,24 | 0,28 |
| Costo | Alto | 0,45 | 0,27 | 0,47 | 0,48 | 0,60 | 0,19 |
| | Bajo | 0,37 | 0,52 | 0,57 | 0,64 | 0,20 | 0,36 |

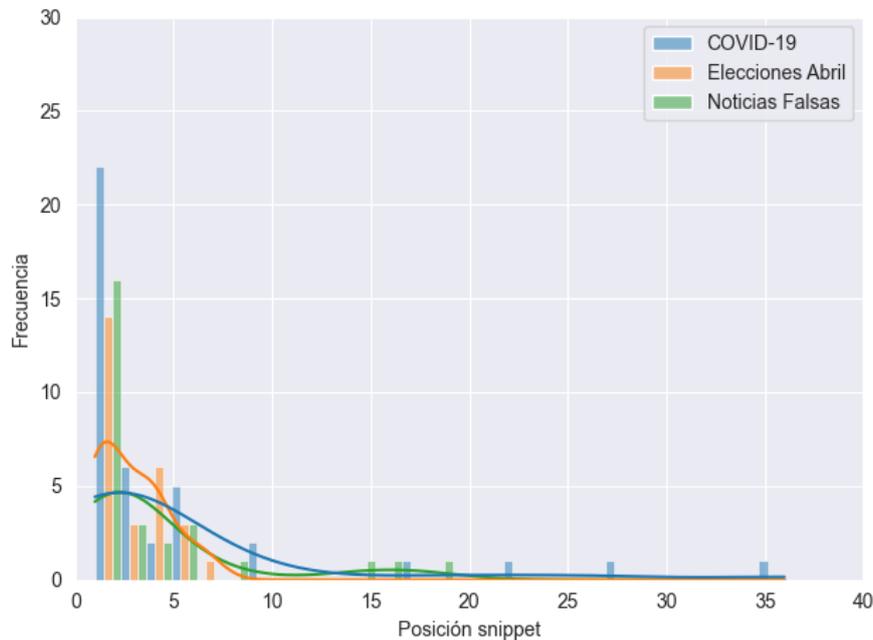
Fuente: Elaboración propia (2024).

3.2. Sesgo de ranking

El sesgo de *ranking* está determinado por la posición en el listado de resultados del *snippet* de la página web guardada por cada participante en la biblioteca personal. En la Figura 2, el eje “x” representa la frecuencia de *snippets* y el eje “y” la posición de cada *snippet* guardada en la biblioteca personal.

Figura 2.

Distribución de la posición de la snippet de la biblioteca personal



Fuente: Elaboración propia (2024).

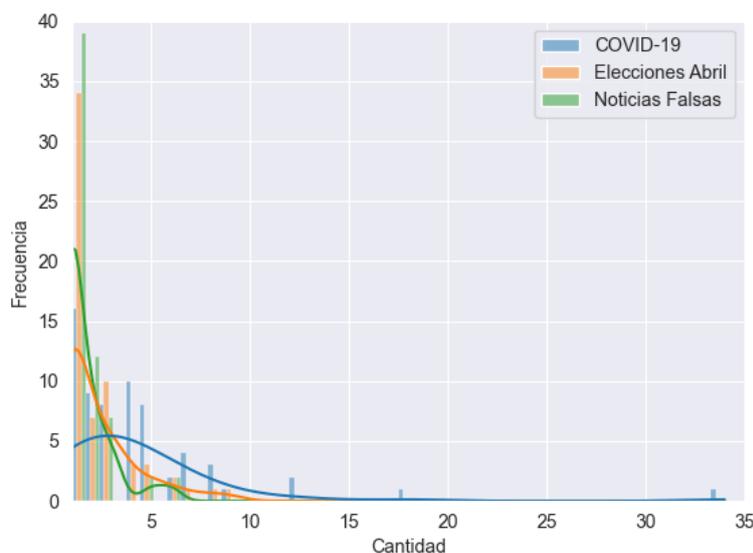
Para la tarea “COVID-19” la distribución de la posición de los *snippets*, si bien se concentra en los primeros resultados, se extiende a lo largo del eje “y” (Figura 2). Lo anterior, muestra el desacuerdo de algunos estudiantes, respecto de los resultados alcanzados mediante sus consultas, lo que se refleja en un aumento en la profundización de la web. En cambio, para las tareas “Elecciones 2021” y “Noticias falsas” la distribución de la posición de las *snippets* seleccionadas se concentra en los primeros lugares del *ranking* de resultados (figura 2). En ambas tareas se observa una mayor operación del sesgo de *ranking*. Donde, la mayoría de las *snippets* guardadas en la biblioteca personal se concentran en la primera posición del *ranking* de resultados.

3.3. Sesgo de disponibilidad

Para representar el sesgo de disponibilidad utilizamos el gráfico de barra que establece la relación entre la cantidad de consultas por persona para cada tarea, demostrando la decisión de exploración de los resultados por parte del estudiantado. En la Figura 3, el eje lateral izquierdo indica el número total de estudiantes y el eje horizontal la cantidad de consultas.

Figura 3.

Cantidad de consultas por tarea



Fuente: Elaboración propia (2024).

En el caso de la tarea “COVID-19” se observa un mayor número de consultas por sobre las otras tareas. Esta tarea muestra que ante resultados diversos y alejados de sus consultas se emiten más consultas. Como casos atípicos dos estudiantes elaboran 18 y 34 consultas, respectivamente (Figura 3). Para la tarea “Elecciones 2021”, se observan menos consultas, pero se mantiene la relación con la similitud entre los términos de las consultas y los resultados provistos por el algoritmo. En esta tarea, un 55% realizaron una única consulta (figura 3). El sesgo de disponibilidad se refuerza en la tarea “Noticias falsas”, con un 63% de participantes que emite una única consulta, con apenas 2 participantes que realizan 6 consultas (figura 3).

4. Discusión

En primer lugar, nos preguntamos cómo operaban los sesgos cognitivos durante una búsqueda de información para resolver una tarea académica, logrando observar y caracterizar los sesgos cognitivos de: anclaje, ranking y de disponibilidad. Se evidencia que la operación de estos sesgos depende de la operación conjunta del pensamiento crítico, la saturación de información y del contexto de las tareas. Se observa que la presencia del sesgo de anclaje principalmente con la consigna en ambas instancias del proceso. Asimismo, el grupo experimental (GE) presenta un menor nivel de anclaje, ya que no advierte la saturación de información, debido a que solo revisan los primeros resultados, lo que se confirma al observar diferencias significativas en la interacción de todas las subdimensiones del EMPC y la saturación sobre el anclaje de la consigna de la tarea y las consultas elaboradas por los participantes. Mientras que solo la subdimensión de utilidad del EMPC en interacción con la saturación presentan diferencias significativas en el nivel de anclaje entre las consultas elaboradas y los *snippets* guardados en la biblioteca personal. Al analizar en detalle las trazas de comportamiento se evidencia que las diferencias de anclaje ocurren en función del contexto de la búsqueda. Se observa que se anclan en los tópicos usados en la consigna de la tarea, aun cuando existan palabras claves que pudiesen conducirlos a una búsqueda más eficiente. Incluso, varios copian y pegan la pregunta en la caja de búsqueda. Dichos términos ancla son utilizados desde la elaboración de la consulta hasta la selección de

fuentes, lo que explicaría que las personas guarden solo los resultados que contengan los tópicos en el *snippet* del resultado. Estos hallazgos validan que la selección de las fuentes de información de la lista de resultados tiene como principal criterio el resumen elaborado por el motor de búsqueda para cada página web.

Respecto del sesgo de *ranking*, se comprueba que el estudiantado está altamente influenciado por la mediación algorítmica. Por lo que se confirma que las personas tienden a seleccionar los primeros resultados del *ranking* (Wu y Kelly, 2015; Schultheiß et al., 2018; Su et al., 2018). Esto se observa en todas las tareas, especialmente en las tareas “Elecciones 2021” y “Noticias falsas”, donde la mayoría de las fuentes guardadas en la biblioteca se ubican en la primera y segunda posición del *ranking*. Por consiguiente, la valoración de una página web como fuente válida de conocimiento está directamente influenciada por la decisión algorítmica. Esto es particularmente riesgoso en el caso de fuentes de información utilizadas para formar opinión y nuevo conocimiento, como en este experimento.

En cuanto al sesgo de disponibilidad se evidencia la tendencia a confiar en la mediación algorítmica si los resultados del buscador contienen fuentes de información similares entre ellas, reduciendo el número de consultas (Ghenai et al., 2020). En los resultados se observa la operación de este sesgo en las tareas “Elecciones 2021” y “Noticias falsas” evidenciado en un menor número de consultas. En contraste, para la tarea “COVID-19” se advierte una reducción de este sesgo, con un mayor número de consultas y/o reformulaciones, con casos atípicos de 12, 18 y 34 consultas. En esta tarea, la decisión de continuar la búsqueda podría obedecer a las diferencias de los datos de las fuentes. Esta temática en particular fue objeto de debate en varios países debido a las diferencias de cálculo de defunciones por COVID-19 (Baeza-Yates, 2021; The Economist, 2021). Por lo que las personas participantes recibieron distintas cifras, incluidas las oficiales. Lo anterior, marcaría la tendencia de aumentar el número de consultas en esta tarea, debido a que intentaron encontrar un consenso en la información. Esta estrategia podría considerarse como un criterio fiable, excepto si la temática de búsqueda es objetivo de campañas de desinformación. Estos resultados confirman lo descrito en la literatura (Ghenai et al., 2020), respecto de la tendencia de las personas a aceptar una lista de resultados homogéneos, ya que los participantes tienden a confiar en los resultados del buscador cuando las fuentes de información son similares entre ellas, tanto en el título como en el *snippet* y con los términos utilizados en las consultas que elaboraron.

En segundo lugar, nos preguntamos si la motivación a pensar críticamente influenciaba significativamente la operación de sesgos cognitivos. En el sesgo de anclaje las variables saturación y las subdimensiones del EMPC muestran una influencia significativa sobre el anclaje entre la consigna de la tarea y las consultas elaboradas por los participantes. La subdimensión de utilidad del EMPC aparece como relevante en el anclaje de las fuentes seleccionadas por los participantes. En general, en la elaboración de consultas se observa un mayor sesgo de anclaje en estudiantes con resultados más altos en la escala EMPC. Sin embargo, se evidencia un aumento en el nivel de anclaje en la tarea “Elecciones 2021”, lo que estaría influenciado por el bajo conocimiento previo de la tarea. Mientras que las respuestas de las personas con bajo EMPC no estarían directamente vinculadas con los resultados alcanzados en la lista de resultados, sino que corresponden a otras fuentes de información tales como el conocimiento previo.

Indagamos si las personas que reciben una lista de resultados saturada manifiestan un comportamiento de búsqueda más sesgado. En este sentido, se determina que, en las fuentes de información guardadas en la biblioteca, se observan diferencias significativas en los niveles de sesgo de anclaje en las diferentes tareas en función de la utilidad y saturación. Sin embargo, la selección de las fuentes se realiza basándose preferentemente en el resumen del resultado, por lo que el nivel de anclaje es mayor en esta relación (*consulta-snippet*). Sin embargo, se observan diferencias en los niveles de anclaje dependiendo del contexto de la tarea en ambos grupos (GC y GE) sobre el nivel de saturación. En concreto, las personas que recibieron una lista de resultados saturada presentan un menor nivel de anclaje respecto de quienes recibieron una lista no saturada, pero tienden a confiar más en los resultados entregados en las primeras posiciones y a emitir menos consultas. Esto puede obedecer a la confianza de las personas en los motores de búsqueda como curadores de información (Noble, 2018). Según esto, las personas no advierten la saturación de información, debido a que solo revisan los primeros resultados y no exploran la lista completa.

5. Conclusiones

La gran cantidad de información disponible en internet nos brinda la oportunidad de resolver problemas de manera más eficiente, siempre que contemos con las habilidades digitales apropiadas para explorar y seleccionar las fuentes más pertinentes para nuestras necesidades. Sin embargo, esta inmensidad de fuentes se convierte en un desafío ante la saturación de información, obligando a los usuarios a utilizar heurísticas o atajos cognitivos que les permitan avanzar más rápido en la resolución de tareas. De esta manera, se comienzan a configurar sesgos cognitivos que podrían llevarnos por rincones de la web que podrían alejarnos de nuestros objetivos e, incluso, sin ser conscientes de ello. En esta investigación estudiamos este comportamiento de búsqueda en estudiantes universitarios chilenos, determinando la presencia de sesgos cognitivos y la influencia de la saturación y la motivación a pensar críticamente durante la resolución de tareas de información en diferentes ámbitos. Los resultados nos muestran la interacción entre las variables independientes sobre la operación de estos sesgos, indicando que los estudiantes tienen a anclarse en la consigna de las tareas, independiente del nivel de saturación al que estaban sometidos. Mientras que el ámbito de la tarea aparece como una condicionante sobre las decisiones cognitivas de los participantes, llevándolos a generar exploraciones más profundas en la lista de resultados generadas por la mediación algorítmica. Por lo tanto, los resultados constituyen una línea de investigación que releva la importancia de las habilidades de búsqueda y selección cuidadosa de información, sobre todo considerando que su utilidad va más allá de la educación formal. Es más, la selección de información que dé garantía de calidad y credibilidad es una competencia clave para la participación democrática y el ejercicio activo de la ciudadanía global. Sin embargo, es necesario destacar que el presente estudio se realiza con estudiantes universitarios de primer ingreso, por lo que los participantes pueden representar un comportamiento pasivo en comparación a estudiantes de cursos superiores. Sería desafiante proponer en un futuro estudios longitudinales. Por otra parte, los participantes resolvieron tres tareas de búsquedas de información en una sesión de trabajo, lo que podría no ser un comportamiento natural para un proceso de búsqueda de información. Por este motivo sería adecuado diseñar el estudio en varias sesiones de trabajo.

6. Referencias

- Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM*, 61(6), 54-61. <https://doi.org/10.1145/3209581>
- Baeza-Yates, R. (18 de marzo de 2021). Estimación del exceso de muertes y fallecimientos por Covid-19 en 2020. *Ciper*. <https://cuwk.short.gy/ciper-covid>
- Bhaskar, M. (2018). *Curaduría: El Poder de la Selección en un Mundo de Excesos*. Fondo de Cultura Económica.
- Buckingham, D. (2008). *Más Allá de la Tecnología. Aprendizaje Infantil en la Era de la Cultura Digital*. Manantial.
- Burbules, N. C. (2012). El aprendizaje ubicuo y el futuro de la enseñanza. *Encounters on Education*, 13(212), 3-14. <https://doi.org/10.24908/eoe-ese-rse.v13i0.4472>
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951715622512>
- Campbell, D. T. y Stanley, J. C. (1995). *Diseños Experimentales y Cuasi-Experimentales en la Investigación social*. En M. Kitaigorodzki (Trad.) (1996). *Amorrortu*.
- Cortada de Kohan, N. (2008). Los sesgos cognitivos en la toma de decisiones. *International Journal of Psychological Research* 2008, 1(1), 68-73. <https://doi.org/10.21500/20112084.968>
- Chander, A. (2016). The racist algorithm. *Michigan Law Review*, 115, 1023. <http://repository.law.umich.edu/mlr/vol115/iss6/13>.
- Faulkner, S. L. y Trotter, S. P. (2017). Theoretical saturation. En J. Matthes (Ed.). *The International Encyclopedia of Communication Research Methods* (pp. 1-2). Wiley.
- Flick, U. (2007). *Introducción a la investigación cualitativa*. Morata.
- Ghenai, A., Smucker, M. D. y Clarke, C. L. (2020, March). A think-aloud study to understand factors affecting online health search. En *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (273-282). <https://doi.org/10.1145/3343413.3377961>
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2018). *Metodología de la Investigación*. McGraw-Hill Interamericana.
- Hernández, D. J., Sánchez, P. M. y Giménez, F. S. S. (2021). La competencia digital docente, una revisión sistemática de los modelos más utilizados. *RIITE Revista Interuniversitaria de Investigación en Tecnología Educativa*, 105-120. <https://doi.org/10.6018/riite.472351>
- Kahneman, D. (2012). *Pensar Rápido, Pensar Despacio*. Debate.
- Kelly, D. y Azzopardi, L. (2015). How many results per page? A study of SERP size, search behavior and user experience. En *Proceedings of the 38th International ACM SIGIR*

- Conference on Research and Development in Information Retrieval* (pp. 183-192). <https://doi.org/10.1145/2766462.2767732>
- Liu, J. (2019). A reference-dependent model of search evaluation. En *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (pp. 405-408). <https://doi.org/10.1145/3295750.3298970>
- Matute, H. (2019). *Nuestra Mente nos Engaña. Sesgos y Errores Cognitivos que Todos Cometemos*. Shackleton books.
- Mesquita-Romero, W. A., Fernández-Morante, C. y Cebreiro-López, B. (2022). Alfabetización mediática crítica para mejorar la competencia del alumnado. *Comunicar: Revista Científica Iberoamericana de Comunicación y Educación*, 70, 47-57. <https://doi.org/10.3916/C70-2022-04>
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- Novin, A. y Meyers, E. (2017). Making sense of conflicting science information: Exploring bias in the search engine result page. En *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (pp. 175-184). <https://doi.org/10.1145/3020165.3020185>
- Olivares-Rodríguez, C., Guenaga, M. y Garaizar, P. (2017). Automatic assessment of creativity in heuristic problem-solving based on query diversity. *Dyna*, 92(4), 449-455. <http://dx.doi.org/10.6036/8243>
- Olivares-Rodríguez, C., Valdés-León, G. y Vidal-Sepúlveda, M. (2021). Influencia de la agenda mediática en las búsquedas de información: una experiencia con estudiantes universitarios. *ReHuSo*, 6(2), 17-28. <https://doi.org/10.33936/rehuso.v6i2.3423>
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Harvard University Press.
- Pereira, S., Fillol, J. y Moura, P. (2019). El aprendizaje de los jóvenes con medios digitales fuera de la escuela: De lo informal a lo formal. *Comunicar*, 27(58), 41-50. <https://doi.org/10.3916/C58-2019-04>
- Pick S. y López A. (1998). *Cómo Investigar en Ciencias Sociales*. Trillas.
- Pothirattanachaikul, S., Yamamoto, T., Yamamoto, Y. y Yoshikawa, M. (2020). Analyzing the effects of "people also ask" on search behaviors and beliefs. En *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (pp. 101-110).
- Roetzel, P. G. (2019). Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business Research*, 12(2), 479-522. <https://doi.org/10.1007/s40685-018-0069-z>
- Sales, D., Cuevas-Cerveró, A. y Gómez-Hernández, J. (2020). Perspectives on the information and digital competence of Social Sciences students and faculty before and during lockdown due to Covid-19. *El Profesional de la Información*, 29(4),1-21. <https://doi.org/10.3145/epi.2020.jul.23>

- Schultheiß, S., Sünkler, S. y Lewandowski, D. (2018). We still trust in Google, but less than 10 years ago: an eye-tracking study. *Information Research: An International Electronic Journal*, 23(3), 799. <http://informationr.net/ir/23-3/paper799.html>
- Simon, H. A. (1957). *Models of Man; Social and Rational*. Wiley.
- Su, N., He, J., Liu, Y., Zhang, M. y Ma, S. (2018). User intent, behaviour, and perceived satisfaction in product search. En *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 547-555). <https://doi.org/10.1145/3159652.3159714>
- Suárez Guerrero, C. y Lloret Català, M. D. C. (2022). La Digitalización de la Educación en Pandemia. Mirada del docente peruano. *REICE-Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 20(4), 125-144. <https://doi.org/10.15366/reice2022.20.4.007>
- The Economist. (2021, 20 de octubre). Tracking COVID-19 excess deaths across countries. In many parts of the world, official death tolls undercount the total number of fatalities. *The economist*. <https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker>
- Tversky, A. y Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293-315. <https://doi.org/10.1037/0033-295X.90.4.293>
- Urgo, K., Arguello, J. y Capra, R. (2019). Anderson and krathwohl's two-dimensional taxonomy applied to task creation and learning assessment. En *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval* (pp. 117-124). <https://doi.org/10.1145/3341981.3344226>
- Vidal-Sepúlveda, M., Olivares-Rodríguez, C. y Cárcamo-Ulloa, L. (2021). What kind of sources do I need? Critical search for information on the Web. *Information Research*, 26(4). <https://doi.org/10.47989/irpaper913>
- Valenzuela, J. y Nieto, A. M. (2008). Motivación y Pensamiento Crítico: Aportes para el estudio de esta relación. *REME*, 9(28). <http://reme.uji.es/articulos/numero28/article3/article3.pdf>
- Wu, W. C. y Kelly, D. (2015). Online search stopping behaviors: An investigation of query abandonment and task stopping. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-10. <https://doi.org/10.1002/meet.2014.14505101030>
- Yagci, N., Sünkler, S., Häußler, H. y Lewandowski, D. (2022). A comparison of source distribution and result overlap in web search engines. *arXiv preprint arXiv:2207.07330*. <https://doi.org/10.48550/arXiv.2207.07330>

CONTRIBUCIONES DE AUTORES, FINANCIACIÓN Y AGRADECIMIENTOS

Contribuciones de los autores:

Conceptualización: Vidal-Sepúlveda, Martha; **Software:** Olivares-Rodríguez, Cristian
Validación: Vidal-Sepúlveda, Martha **Análisis formal:** Vidal-Sepúlveda, Martha; **Curación de datos:** Vidal-Sepúlveda, Martha; Olivares-Rodríguez, Cristian; **Redacción-Preparación del borrador original:** Vidal-Sepúlveda, Martha **Redacción-Revisión y Edición:** Vidal-Sepúlveda, Martha; Olivares-Rodríguez, Cristian **Visualización:** Olivares-Rodríguez, Cristian **Supervisión:** Olivares-Rodríguez, Cristian **Administración de proyectos:** Vidal-Sepúlveda, Martha **Todos los/as autores/as han leído y aceptado la versión publicada del manuscrito:** Vidal-Sepúlveda, Martha; Olivares-Rodríguez, Cristian

Financiación: Esta investigación no recibió financiamiento externo.

Conflicto de intereses: Los autores declaran no tener ningún conflicto de interés.

AUTORES:

Martha Vidal-Sepúlveda

Universidad Central de Chile, Chile.

Profesora de Historia y Geografía (1999) por la Universidad de Concepción, Magíster en Desarrollo Curricular y Proyectos Educativos (2014) por la Universidad Andres Bello y Doctora en Comunicación (2023) de la Universidad Austral de Chile y la Universidad de La Frontera. Desarrollo estudios sobre Alfabetización digital, Ciudadanía digital y las problemáticas educomunicativas que emergen en el proceso de resolución de problemas en contextos educativos. Asimismo, analizo el comportamiento de consumo de medios digitales a través de la búsqueda de información y la influencia del pensamiento crítico en la conducta de los lectores jóvenes de noticias. Finalmente, llevo a cabo estudios sobre competencias digitales docentes ante a la aparición de la Inteligencia Artificial Generativa.

martha.vidal@uccentral.cl

Índice H: 3

Orcid ID: <https://orcid.org/0000-0002-0929-8179>

Cristian Olivares-Rodríguez

Universidad Alberto Hurtado, Chile.

Ingeniero Civil Informático (2006) por la UCSC (Chile), Máster en visión por computadora e inteligencia artificial (2010) por la UAB (España) y Doctor en Ingeniería (2017) por la Universidad de Deusto (España). Académico en ciencia de datos e ingeniería de software. Estudio modelos de uso, apropiación y coproducción de tecnologías y, en particular, de la Inteligencia Artificial, es un área interdisciplinaria que combina elementos de la sociología, la tecnología, la comunicación y la gestión del conocimiento. Aquí se examinan cómo las tecnologías son adoptadas, adaptadas y transformadas por las personas y las comunidades en diversos contextos, particularmente en relación con la gestión y desarrollo de los territorios, considerando que las tecnologías son productos de procesos sociales, pero también los configuran.

colivares@uahurtado.cl

Índice H: 8

Orcid ID: <https://orcid.org/0000-0002-4991-5784>